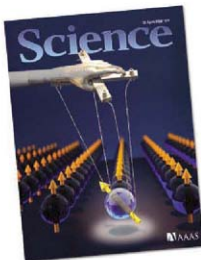# Science

COVER

Optical or electrical pulses can rapidly locate and manipulate the spin of a single electron in a quantum dot or of a nitrogen vacancy in diamond. Such techniques represent progress toward solid-state quantum computing (see pages 349 and 352).

Image: Peter Allen and Jesse Berezovsky

EDITORIAL

306

319

Science

328 & 346

REPORTS *CONTINUED...*

379

△AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Financial trading can wreak havoc on physiology.

SCIENCE**NOW**
www.sciencenow.org DAILY NEWS COVERAGE

Living in a World of Unfamiliar Voices
A woman with a lifelong inability to recognize voices is a medical mystery.

Bad Day for the Dow? Blame Hormones
Market activity is reflected in financial traders' testosterone and cortisol levels.

Case Closed for Free Will?
The unconscious brain makes choices several seconds before the conscious mind knows about it.

Modeling cell signaling.

SCIENCE **SIGNALING**
www.sciencesignaling.org
THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

ST NETWATCH: BioModels Database
Explore this collection of mathematical models of biological pathways; in Modeling Tools.

ST NETWATCH: CellDesigner
Build mathematical models of biological pathways with CellDesigner; in Modeling Tools.

ST NETWATCH: Systems Biology Workbench
Integrate mathematical models and share data between different applications; in Modeling Tools.

Help make public policy.

SCIENCE **CAREERS**
www.sciencecareers.org/career_development
CAREER RESOURCES FOR SCIENTISTS

A Matter of Policy
B. Vastag
Fellowships are available that let scientists contribute to local and national policies. >> *Editorial p. 289*

Tooling Up: On Headhunters
D. Jensen
Recruiters may seem to offer some advantages, but early-career scientists should still keep their heads.

Home Stretch to Graduation
E. Pain
Submitting your Ph.D. dissertation isn't always a relief.

From the Archives: What's Love Got to Do With It?
I. S. Levine
Our Mind Matters expert studies the pros and cons of scientists tying the knot.

SCIENCE **PODCAST**

Download the 18 April *Science* Podcast to hear about seasonal influenza, aquatic mercury in terrestrial food webs, making synthetic fuels, a geometric model of music, and more.
www.sciencemag.org/about/podcast.dtl

Separate individual or institutional subscriptions to these products may be required for full-text access.

<< Acid Tests

One worrying consequence of increasing atmospheric carbon dioxide is ocean acidification. As the oceans absorb carbon dioxide, the pH will fall, making it more difficult for calcifying organisms such as corals to produce and maintain their skeletons. This change could have disastrous consequences for many types of marine life, and also serious repercussions for terrestrial species including humans. Now **Iglesias-Rodriguez et al.** (p. 336) report that, contrary to expectations, high atmospheric carbon dioxide levels may actually increase calcification by the coccolithophore species *Emiliania huxleyi*. Thus, the ecological and biogeochemical effects of rising atmospheric carbon dioxide levels may not be straightforward.
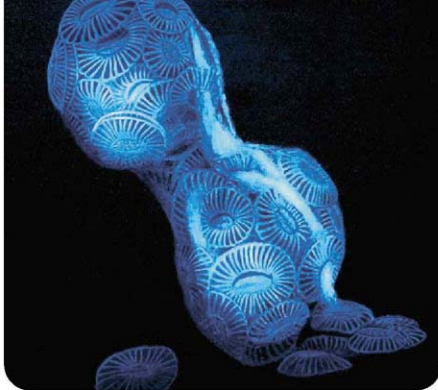
Stocking the Malaria Arsenal

An unassuming plant from China shows promise as a source of powerful antimalarial drugs. Artemisinin and its various derivatives, often delivered in combination with preexisting antimalarial drugs, are proving tough in the fight against malaria. **White** (p. 330) reviews the derivation of the artemisinin arsenal, its current deployment, and how combination artemisinin therapies fit in with global policy initiatives to put malaria to rest.

One-Way Ticket for Influenza

During the past 5 years, influenza epidemics have been seeded by viruses that originated in east and Southeast Asia, probably from temporally overlapping epidemics rather than by persistence. **Russell et al.** (p. 340) have taken 13,000 isolates of influenza virus and analyzed the HA1 domain of the hemagglutinin to investigate the ancestry of the circulating strains. Travel and trade connections explain the global dissemination of influenza strains on a one-way route out of Asia, taking about 6 to 9 months to reach Europe and North America. Several months later, these strains arrive at their evolutionary graveyard in South America. Thus, the antigenic characteristics of currently circulating viruses in east and Southeast Asia may be key to forecasting vaccine needs.

Math and Music

Musicians and composers use a variety of techniques for grouping notes that reflect an intuitive sense of their relatedness. For example, major and minor triads (chords of three notes) are considered to be similar, even though they do not sound exactly identical. Although such intuitions are common, they are difficult to quantify. **Callender et al.** (p. 346; see the Perspective by **Hall**) use the methods of geometry to model how western musicians traditionally classify pitch sequences. Geometrical spaces exist in which chord types can be represented, and distance between chords in the mathematical sense corresponds to similarity in the intuitive sense. This translation of musical theory into the language of contemporary geometry provides a framework for discerning relationships among musical works and tools for interpreting compositions.

Solid-State Spin Control

The coherent control of a single spin on a quantum dot is a fundamental requirement for solid-state quantum information processing. **Berezovsky et al.** (p. 349) demonstrate such coherent manipulation of a single electron spin on a quantum dot using ultrafast optical pulses. Using the optical Stark effect and a series of ultrafast optical pulses, they rotate the single spin through arbitrary angles up to π radians on picosecond time scales. The observed spin rotations constitute true single-qubit operations, performed on a time scale much shorter than the coherence time, and are readily scalable to the large number of operations needed for practical applications. However, in the case of
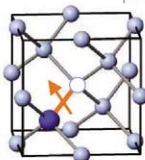
the nitrogen vacancy (NV) in diamond, the single spin to be manipulated sits in a bath of background spins, which can be detrimental to the quantum dynamics of the NV center. **Hanson et al.** (p. 352, published online 13 March 2008; cover) present experimental and numerical modeling aimed at understanding the coupling and eventually controlling the coherence of the quantum dynamics of the spin of the NV center.

Cellular Orienteering

How do cells control membrane retraction and the direction of cell movement? **Witze et al.** (p. 365; see the Perspective by **Bowerman**) examined cells responding to the developmental signaling protein Wnt. Cultured human melanoma cells responded to Wnt5A by accumulating a cluster of receptor, adhesion, cytoskeletal, and motor proteins near the cell surface. When cells were orienting in a gradient of chemokine concentration, the protein cluster was localized asymmetrically at the posterior of the cell. This structure may help cells integrate the actions of receptors that mediate cell adhesion and cell signaling with cytoskeletal components to control membrane retraction and the direction of cell movement.

The Yin and Yang of Neuronal Maintenance

During development, more peripheral neurons project to target organs than are ultimately needed. The neurons then compete for neurotrophic factors that are secreted by target cells. **Deppmann et al.** (p. 369, published online 6

March 2008) now explain how some neurons manage to survive, whereas others die, even though they have similar access to sustaining growth factors. The answer appears to depend on a series of feedback loops. Nerve growth factor (NGF) secreted by target cells not only stimulates expression of its own receptor, but also promotes expression of other factors that can cause neuronal cell death. The surviving neurons appear to have sufficiently strong NGF signaling to withstand the antagonistic signals.

Improving on Immunoglobulins

Intravenous immunoglobulin (IVIG) therapy uses pooled fractions of human serum immunoglobulin G (IgG) to treat a variety of conditions, including autoimmune diseases. The treatment relies in some way on the anti-inflammatory activity of a subfraction of the immunoglobulins applied, and sialyation of N-linked sugars in the constant portion of the IgG chains is known to be important. Now, **Anthony et al.** (p. 373) define a specific sialic acid–galactose linkage required for anti-inflammatory activity. A recombinant sialylated IgG Fc fragment could recapitulate the anti-inflammatory activity of IVIG, suggesting that it might be possible to capture the effectiveness of IVIG without the need for human donors.

Reconstitution of Bacterial Pilus Assembly

Adhesive pili are filamentous protein complexes on bacterial surfaces, which mediate the adhesion of pathogenic bacteria to host tissues. Pili serve as a paradigm for studying ordered macromolecular assembly reactions at the bacterial cell membrane. **Nishiyama et al.** (p. 376, published online 27 March 2008) now describe the complete in vitro reconstitution of an assembly and secretion system for adhesive pili from purified pilus proteins, using type 1 pili from uropathogenic *Escherichia coli*. The reconstitution reveals how a protein catalyst can accelerate the ordered assembly of a supramolecular protein complex.

Receptor-dsRNA-Receptor

Toll-like receptors recognize molecules associated with pathogens and initiate inflammatory responses. For example, Toll-like receptor 3 (TLR3) recognizes double-stranded RNA (dsRNA), an intermediate in viral replication. The TLR3 ectodomain binds as a dimer to dsRNA, but the molecular basis for signaling remains unclear. **Liu et al.** (p. 379) now report the structure of a complex between two mouse TLR3– ectodomains and dsRNA. Two horseshoe-shaped TLR3-ectodomain monomers bind to opposite faces of the dsRNA through their N- and C-termini and dimerize through their C-termini so that the N-termini are at opposite ends of the linear dsRNA molecule. This dimerization mode could mediate signal transduction by facilitating dimerization of the receptor cytoplasmic domains.

Little Change, Large Consequence

Occasionally, small changes in sequence change the overall architecture of large protein assemblies. One system for understanding protein assembly is the bacterial flagellar filament, the prototype of which comes from *Salmonella* and contains eleven protofilaments. **Galkin et al.** (p. 382), now show that the homologous flagellar filament of *Camylobacter* contains only seven protofilaments. The difference may be related to sequence divergence in a region of flagellin that in *Salmonella* is involved in coiled-coil formation and is recognized by the vertebrate Toll-like receptor 5 (TLR5). *Campylobacter* is not recognized by TLR5, and its evasion may have driven the change in quaternary structure.

Antidepressants and Adult Brain Plasticity

The mechanism of action of antidepressant drugs is still unclear, but neuronal plasticity may be important. **Maya Vetencourt et al.** (p. 385) investigated whether chronic treatment with antidepressants restores plasticity in the adult visual system of the rat. The authors used two classical models of plasticity, the ocular dominance shift of visual cortical neurons following monocular deprivation and the recovery of visual function in the adult after long-term monocular deprivation. Surprisingly, chronic administration of antidepressants increased brain-derived neurotrophic factor expression in the visual cortex and reduced intracortical inhibition, thus restoring ocular dominance plasticity in adulthood and promoting the recovery of vision in adult rats. Antidepressants may thus increase plasticity throughout the brain, potentially explaining their antidepressant effects.

New Career Paths for Scientists

Bruce Alberts is Editor-in-Chief of *Science*.

Last week on this page, I stressed the benefits to both science and society of transitioning well-trained scientists into a broad array of endeavors, in research and in other roles. Here I suggest two strategies that could help achieve this goal.

For more than 30 years, the Science and Technology Policy Fellowships of the American Association for the Advancement of Science have recruited U.S. scientists and engineers at various stages of their careers, from ages 25 to 72, to work in the U.S. federal government for a year. Similar 10-week fellowships at the U.S. National Academies in Washington, DC, allow graduate students and postdoctoral fellows in science and engineering from many nations to contribute to science and technology policy issues. Other fellowship opportunities in the United States and elsewhere provide exposure to the worlds of policy-making, teaching, and communication, among others (see *Science* Careers, p. 390).

These valuable programs serve multiple purposes, most obviously allowing scientists and engineers to explore possible careers outside of academia and industry. After fellows complete such programs, they return home and share their experiences. Thus, a single fellow can provide an entire academic department with a broader view of career paths. Working with a science fellow can also make an organization or government agency aware of the advantages of hiring full-time staff with scientific talents and connections, permanently increasing its scientific capabilities.

The several thousand past participants in these fellowship programs are engaged in various pursuits. Many are research scientists, but others have entered careers in policy, science education, journalism, and environmental protection, among others. Exemplars include physicists Rush Holt, a U.S. congressman, and E. William Colglazier, the chief of staff at the National Academies. Scientists in such nontraditional careers are invaluable as two-way interpreters: people who can neatly bring the benefits of scientific analysis to their institution or profession, as well as help traditional scientists better understand how their science might contribute in new ways. Even a single such individual can make a huge difference.

There would therefore be many advantages to expanding these types of opportunities. Perhaps the simplest way would be through new short-term programs that allow fellows to sample a career in government, pre-college education, nongovernmental organizations, the media, or industry. If offered for a period of 4 months or so, such fellowships could be accommodated as temporary excursions from traditional career paths. Minimal stipends could cover living costs, as is done for some other fellowship programs, with groups of fellows being mentored by professional staff.

More ambitious would be a new type of graduate program for scientists, with a branching set of options after the first or second year. Although many students would continue to pursue the standard research path, other options would specifically prepare a student to become a professional policy analyst, a science education researcher, a science-oriented journalist, or a science curriculum specialist in a school district, for example. Although there are good stand-alone programs leading to some of these careers, most remain unconnected to standard science Ph.D. programs and are of limited capacity.

Developing new, integrated programs will require that partnerships be forged with other organizations; these could be established regionally to involve students from different universities. They can be viewed as the next logical step from programs such as the U.S. National Science Foundation (NSF) Graduate Teaching Fellows, which enables graduate students in science, technology, engineering, or math to broaden their training through extensive interactions with young students and their teachers.

According to NSF, there are more than 45,000 postdoctoral fellows in the natural and social sciences in the United States alone. Many will gather in Boston next week at the National Postdoctoral Association's Annual Meeting. They and the tens of thousands of graduate students just behind them in the pipeline represent a tremendous resource for the future. It is good news that a surprising percentage are interested in using their science training in nontraditional ways. Those of us who are their mentors must help them do so. — **Bruce Alberts**
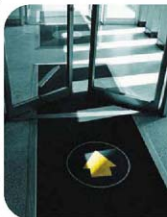
BEHAVIOR
Crabbing in the Chesapeake

As the state crustacean of Maryland, the blue crab (*Callinectes sapidus*) is dear to the hearts and stomachs (and wallets) of many residents. The commercial crabbing season opened just a few weeks ago, and pessimistic forecasts for the size of this year's harvest have led to draft proposals that more stringent protection of adult females be implemented. Kamio et al. describe a laboratory study of the courtship behavior of the male blue crab and interpret it as an adaptation to their environment. They observed that upon being presented with an inaccessible female (corralled by plastic mesh), male crabs spread their chelae (the first pair of legs in a decapod crustacean, commonly referred to as claws in crabs and lobsters) and elevated their swimming legs (the last pair of legs) into a dorso-medial position. When in this stance, the males paddled their swimming legs circularly and 180° out of phase. Particle imaging velocimetry revealed a forward-directed flow of water traveling at an average speed of 3 cm s⁻¹. The authors suggest that the marshy habitat of the Chesapeake Bay allows female crabs to avoid predators while signaling their presence by pheromone; the males have adopted courtship paddling as a means of wafting their own chemical lures toward unseen females in order to coax them into open water. — GJC

J. Exp. Biol. **211**, 1243 (2008).

PHYSICS
Probing Non-Standard Charges

The Standard Model of particle physics describes three of the four natural forces, detailing the strengths of interactions among the protons, electrons, and neutrons that make up atoms, as well as the family of quarks that in turn make up these particles. The framework gives rise to charge quantization, as measured in discrete units of electron charge e, and also the charge neutrality of atoms. However, the violation of certain symmetries that describe the underlying physics of all these forces—as is evident from the dominance of matter over antimatter in the universe—indicates that there is something happening beyond the Standard Model. Much theoretical and experimental

effort is directed toward exploring this regime, with some work suggesting that the notion of charge quantization and charge neutrality should be abandoned. In efforts to detect the minis-cule charges that could explain the broken symmetries, Arvanitaki et al. describe a sensitive method for detecting charge on an atom based on the interference of atom waves. They argue that during splitting and recombination of a con-densed cloud of rubidium atoms, a departure from charge neutrality of as little as 10^{-28} e should be detectable as a phase shift in the interfering atom waves. This cold atom approach may provide an alternate route to looking beyond the Standard Model. — ISO

Phys. Rev. Lett. **100**, 120407 (2008).

CHEMISTRY
Taking the Heat

Ceramics are often prepared by heating struc-turally well-defined molecular or polymeric pre-cursors to very high temperatures; in the process, peripheral hydrogen and halide (X) atoms are expelled as HX gases, and the severed bonds of the remaining solid rearrange into a rigid network that lends the material stability. When prepared in this manner, the amorphous ceramic SiBNC shows impressive resistance against thermal or oxidative degradation. Somewhat surprising is the key con-tribution of carbon, which traditionally segregates from nitrogen during pyrolysis. Sehlleier et al. monitored the fate of the carbon precursor (methylamine) during the synthesis of this mate-rial, and found strong evidence that in this case, C-N bonds remain intact even at 1400°C. Their study relied on double isotopic labeling (^{13}C and ^{15}N) of the precursor and subsequent analysis of the ceramic by solid-state nuclear magnetic reso-nance spectroscopy. Application of the rotational echo double-resonance technique revealed the distribution of bonding distances between the two labeled nuclei and suggested some degree of multiple bonding through π overlap. — JSY

Angew. Chem. Int. Ed. **47**, 10.1002/anie.200705786 (2008).

CELL BIOLOGY
Divide and Rule

Intracellular cholesterol metabolism is controlled by sterol regulatory element–binding proteins (SREBPs), which are stored in an inactive mem-brane-bound form within the endoplasmic reticu-lum (ER) compartment via an interaction with ER-restricted sterol sensors. When the level of cellular sterols is depleted, the sensors let go of the SREBPs, which are then trans-

Colocalization (yellow) of S1P in the Golgi.
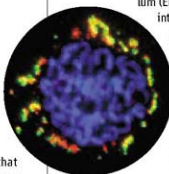
ported to the Golgi complex, where they are cleaved by site-1 protease (S1P). This cleavage liberates the transcription factor domain from the membrane, and it then enters the nucleus and activates the expression of genes involved in sterol and phospholipid metabolism. During inter-phase, like many Golgi proteins, S1P cycles

between the ER and the Golgi complex, and when ER and Golgi membranes merge, for example after treatment with the drug brefeldin A, SREBP cleavage is activated. During mitosis, the Golgi disassembles, and some evidence suggests that a fraction of the Golgi and ER membranes merge; so how then can the cell avoid inadvertently activating SREBPs? Bartz *et al.* show that during mitosis, S1P and SREBP reside in separate compartments: SREBP remains trapped in the ER, and S1P is sequestered in the clustered remnants of the disassembled Golgi. — SMH

EMBO J. **27**, 948 (2008).

Epoxide Exfoliants

The graphite form of carbon consists of strongly bonded two-dimensional sheets stacked relatively loosely on top of each other. One approach for separating graphite into these individual graphene sheets or multilayers—a morphology with broadly useful emerging electronic properties—involves preliminary oxidization. Solid-state nuclear magnetic resonance studies suggested that the major oxidation product is an epoxide, similar to the product formed upon oxidizing single-walled carbon nanotubes. Chattopadhyay *et al.* have obtained Raman spectral evidence for epoxide formation on graphite, and could furthermore quantify the extent of the reaction by methyltri-oxorhenium-catalyzed transfer of the oxygen atoms to triphenylphosphine, a reaction pathway specific to epoxides. They also imaged functionalization of individual flakes using high-resolution transmission electron microscopy and scanning tunneling microscopy. The reduced and oxidized forms showed similar basal plane roughness, which suggests that epoxidation occurs mainly at edge sites. — PDS
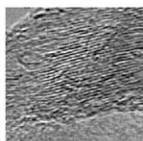
J. Am. Chem. Soc. **130**, 10.1021/ja711063f (2008).

Y not X?

In mammals, a disproportionately high number of genes from the X chromosome have been moved, via transposable elements, to the autosomes. By examining all retrotransposed genes in three placental and one marsupial mammal, Potrzebowski *et al.* have found that only a subset of these genes, the majority originating from the X chromosome, have retained their functions. In mice, most of the relocated genes originating from the X chromosome are specifically expressed in the testes, while those originating from other chromosomes are expressed broadly. The movement is due to a process known as meiotic sex chromosome inactivation, where replicated genes take over the function of their X-linked parents, which are silenced during meiosis. On the basis of these results, the authors hypothesize that the origin of these meiotic retrogenes dates to the divergence of the placental and marsupial mammals. Because meiotic X inactivation reflects the differentiation of the X and Y chromosomes, these chromosomes may be younger than previously suggested. — LMZ

PLoS Biol. **6**, e80 (2008).

Science Signaling
AAAS

<< Sirtuin to Become Astrocytes

Although neural progenitor cells (NPCs) can differentiate into neurons, astrocytes, or oligodendrocytes, brain injury stimulates the production of astrocytes in preference to neurons. Noting that some neurological diseases are associated with inflammation and oxidative conditions, Prozorovski *et al.* investigated the effects of redox state on NPC differentiation. The fraction of cultured NPCs that differentiated into astrocytes was larger under oxidative conditions, whereas the fraction that differentiated into neurons was smaller. Conversely, NPCs cultured under reducing conditions were more likely to differentiate into neurons and less likely to differentiate into astrocytes. The amount of the histone deacetylase Sirtuin1 (Sirt1) was increased in NPCs cultured under oxidative conditions, and *Sirt1* knockdown blocked the effects of oxidation on differentiation. Oxidative conditions promoted the association of Sirt1 with the transcription factor Hes1 and decreased the expression of *Mash1*, which encodes a neurogenic transcription factor. Mouse pups injected with a pro-oxidant showed an increase in Sirt1 in a germinal region of the brain, along with a decrease in neurogenesis; in utero knockdown of *Sirt1* increased neurogenesis in such oxidant-treated pups. Thus, the authors conclude that nontoxic manipulation of redox conditions can influence NPC fate, and that Sirt1 plays a critical role. — EMA

Nat. Cell Biol. **10**, 385 (2008).

On the Hoof

Painters used to depict racing horses with both pairs of legs outstretched. Then in the 1870s, British photographer Eadweard Muybridge became the first to show how horses really run by setting up a row of cameras that snapped as a horse galloped past. In this gait, all four feet leave the ground only as the horse gathers itself for the next stretch—the opposite of what artists imagined. The photos are part of a major exhibit on the horse opening on 17 May at New York City's American Museum of Natural History.

Seeing Stars in Chile

European stargazers have bad and good news. Some European telescopes at the La Silla Observatory will be shut down to save money. But another European observatory in Chile is seeing new kinds of stars—from Hollywood.

At La Silla, a 2.2-meter telescope will be decommissioned after next year, and visiting scientists will no longer have access to the 3.6-meter scope as it moves full-time into a planet-hunting project called HARPS. Some astronomers are grumbling. "Some of the instruments at La Silla cost peanuts to keep operational," says Christiaan Sterken of the Free University of Brussels, Belgium. The telescopes may be small, but they are needed for fields such as asteroseismology, he says, and for training young astronomers.

Meanwhile, as if to rub it in, James Bond actor Daniel Craig visited the observatory at Cerro Paranal, home to the quadruple 8.2-meter Very Large Telescope. A 300-person film crew spent 4 days there shooting scenes for the film *Quantum of Solace*. The spacy-looking residence building was chosen "both for its exceptional design and its remote location in the Atacama Desert," says Michael Wilson, a producer. It is "the perfect hideout for our villain."

Villain's hideout—the residencia at Atacama.

So will proceeds from moviemaking be the answer for telescopes that are down on their luck? Apparently not. Andreas Kaufer, director of the Paranal Observatory, says the film company compensated them for extra expenses "but not much more."

Orcas and Oil

Nearly 2 decades have passed since the *Exxon Valdez* spilled 42 million liters of crude oil into Alaska's Prince William Sound. But the effects are still reverberating among the local killer whales.

Marine biologist Craig Matkin of the North Gulf Oceanic Society in Homer, Alaska, and colleagues collected data—including individual photo IDs—on two genetically distinct whale groups for 5 years before the spill. When it occurred, one group (AT1) was photographed swimming in oil less than 24 hours later; the other (AB) was spotted in oily waters 6 days later.

When killer whales disappear from a pod for more than a year, they are presumed dead. In the first year after the spill, the AB group lost 14 of its 36 members, and its birthrate plummeted. The AT1 group plunged from 22 to 13. "The whales that surfaced in the oil probably suffered irreparable damage to their lungs," says Matkin; others were harmed by eating oily seals and sea lions.

Neither group has recovered, although the AB pod is slowly increasing its numbers, the scientists reported last month in *Marine Ecology Progress Series*. However, the AT1 group has shrunk to seven and "will probably go extinct," says Matkin, because the surviving adults may be too closely related to mate.

It's a sad tale that "a great paper that shows the value of long-term studies," says John Durban, a marine mammalogist at the Center for Whale Research in Friday Harbor, Washington.

AQUATIC COUSINS

Elephants, sea cows, and hyraxes make an odd family photo, but they are one another's closest living relatives. A new study of fossil teeth indicates that, 37 million years ago, at least two members of the elephant clan still spent most of their time in the water.

Analyzing the stable isotope ratios in the tooth enamel of two fossilized elephant cousins from Egypt (see drawing), a team led by paleobiology Ph.D. student Alexander Liu of the University of Oxford in the U.K. found that the carbon and oxygen composition of the teeth was much nearer that of modern-day aquatic mammals than that of terrestrial ones. They reported their findings online last week in the *Proceedings of the National Academy of Sciences*.

The authors "have a powerful set of methods and a powerful framework now," says paleontologist William Sanders of the University of Michigan, Ann Arbor, who says he hopes they will apply the same techniques to reveal the lifestyle of 60-million-year-old fossils that may be ancestral to both elephants and sea cows.

Moeritherium.

CREDITS (TOP TO BOTTOM): AMNH LIBRARY SPECIAL COLLECTIONS; ESO; LUCI RETTENBACH, STONY BROOK UNIVERSITY

<< Slice of Life

A LONG WAIT. In 1964, Peter Higgs proposed a theory to explain why fundamental particles have mass, predicting the existence of a new particle that came to be known as the Higgs boson. Earlier this month, the now-retired physicist ventured out of his home in Edinburgh to visit the machine that he hopes will prove him right: the Large Hadron Collider (LHC) at the CERN particle physics lab near Geneva, Switzerland. Higgs told reporters that he hopes LHC, which goes online this summer, will have confirmed his theory before his 80th birthday in May 2009. If not, he joked, "I'll just have to ask my [general practitioner] to keep me alive a bit longer."

Paul Collier, CERN's group leader for beam operations, says Higgs "was quite impressed by the nitty-gritty of all the material that makes up a particle accelerator." He adds that, amid all the hectic preparations, it's easy to lose sight of the overall goal. "With Peter there, it reminded me why we're doing it."

The $6 billion LHC, built in a 27-kilometer-long tunnel that straddles the Swiss-French border, will smash together beams of protons with enough energy to match the conditions that existed moments after the big bang. Physicists expect to learn more about the fundamental nature of particles and forces, even if they do not find the Higgs boson.

IN PRINT

READING, CRITICALLY. A prominent publisher of U.S. textbooks is reviewing sections on global warming in a public policy textbook after a high school student complained that it ignores the strong scientific consensus on the topic. Matthew LaClair, a senior at Kearney High School in New Jersey, objected to the assertion in the 2006 edition of *American Government: Institutions and Policies* that "science doesn't know whether we are experiencing a dangerous level of global warming or how bad the greenhouse effect is, if it exists at all." The Houghton Mifflin textbook contrasts

concerns of "activist scientists" about global warming with observations by "skeptics."

LaClair tipped off the Center for Inquiry, an Amherst, New York, group that fights for the separation of church and state, which polled James Hansen, director of NASA's Goddard Institute for Space Studies in New York City, and other top climate researchers. The scientists found several statements that were "profoundly mistaken in ways that will mislead students about the facts and science of global warming," Hansen wrote last month in a letter to the publisher. The latest edition of the book, which came out before the center issued a press release about it last week, does not contain the statement questioning the existence of warming. But other sections to which the scientists objected are still in it. "Global warming is virtually unequivocal," says LaClair. "It's vitally important that students get this knowledge."

AWARDS

Stanford University developmental biologist **Philip Beachy** and Harvard University geneticist **Clifford Tabin** have won the March of Dimes Prize in Developmental Biology for their contributions to the understanding of how hedgehog genes guide an embryo's development. They will share $250,000.

Terence Tao, a mathematics professor at the University of California, Los Angeles (UCLA), is the winner of this year's Alan T. Waterman Award from the U.S. National Science Foundation. The prize recognizes outstanding scientists under the age of 35. Tao is 32 and has been on UCLA's faculty for the past 12 years. The 3-year, $500,000 award is in honor of his contributions to the areas of partial differential equations, combinatorics, number theory, and harmonic analysis.

Got a tip for this page? E-mail people@aaas.org

Movers

A GOOD BASELINE. When Hugh "Mont" Montgomery, 60, takes the helm at the Thomas Jefferson National Accelerator Facility (JLab) in Newport News, Virginia, in September, he'll inherit a facility that's doing better than most U.S. Department of Energy national labs. Construction on an upgrade to JLab's main accelera-

tor is scheduled to start this summer, and JLab's 2008 budget has rebounded after some very lean years—which looks good compared with the shutdowns and layoffs at the Fermi National Accelerator Laboratory in Batavia, Illinois, where Montgomery has been associate director for research for the past 6 years.

Looking down the road, Montgomery sees the possibility

of building an electron-ion collider on the 24-year-old lab's 81-hectare campus or perhaps expanding its free-electron laser for materials sciences. More immediately, he's leaving the summer open in case son Richard makes the U.S. Olympic rowing team. "There's a slight chance they'll qualify for Beijing," he says. "And if they do, I'll be there."

CREDITS (TOP TO BOTTOM): CERN; SOURCE: MATTHEW LACLAIR; FERMILAB

NEWS>>

THIS WEEK

Archaeological peace plan

302

Bloodletting at a medical journal

304

CLIMATE CHANGE

IPCC Tunes Up for Its Next Report Aiming for Better, Timely Results

The international team of climate change scientists that produced an influential series of reports last year—and won the 2007 Nobel Peace Prize—will be doing things a little differently in the future. Government delegates to the Intergovernmental Panel on Climate Change (IPCC), meeting last week in Budapest, Hungary, approved a plan for the 20-year, 100-nation enterprise that would generate more precise and relevant information on climate change—without taking any longer than the current 6-year gap between reports. To do so, the delegates endorsed procedural changes that scientists had proposed to streamline the process.

Having heard persuasively in successive IPCC reports that human-induced climate change is real, governments now want more information on what those impacts will be and how the world might begin to curb emissions of greenhouse gases. But scientists say the process used to generate the previous four reports can't deliver the additional detail and greater certainty that policymakers crave in the same 6-year time frame. Rejecting a proposal for an interim report in 4 years and a final report in 8 years, the delegates instead agreed to modify the process itself to achieve the desired results.

The first change would ditch the practice of prescribing the scenarios of economic and technological progress driving future greenhouse gas emissions that researchers should incorporate into their modeling, the first step in the process. Delegates also backed the idea of having the communities that correspond to the panel's three working groups on the science of climate change, its impacts, and mitigation strategies develop their studies in parallel rather than sequentially. Scientists say these changes will reduce the level of uncertainty in their findings, deliver more regional details, and provide policymakers with better clues on how to curb climate change—without length-ening the time from start to finish. The new regime "will expedite and improve the process," says Richard Moss of the World Wildlife Fund, who helped coordinate the effort for IPCC.

"We're going along with the community," says Harlan Watson, head of the U.S. delegation, explaining that the new procedures will "let the science drive the process." Scientists who feared more drastic measures seem pleased with the tweaks to the system. "The overarching sense was if it ain't broke don't fix it," says ecosystem modeler Kathy Hibbard of the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, who attended as an observer. The new structure lays out

Homing in. Scientists hope to improve the geographic resolution of global climate models, shown here in old (left) and new models from the National Center for Atmospheric Research.
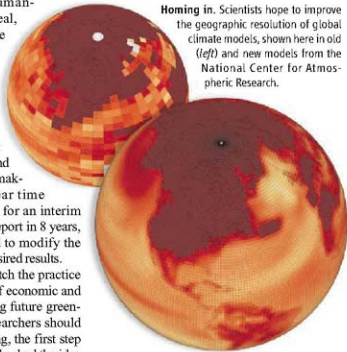
"a rigorous and deliberate approach, taking the time required," says NCAR atmospheric chemist and U.S. delegate Susan Solomon.

The 40 emissions scenarios used in modeling studies cited by last year's IPCC reports employed outdated assumptions about population growth, energy use, and emerging emissions-reducing strategies, experts say. "For example, nobody thought seriously about stor-ing carbon under the ground," says Tom Kram of the Netherlands Environmental Assessment Agency. To correct that problem, IPCC sponsored a series of workshops over the past 2 years in which the community chose four scenarios from published papers that span a range of possible economic and technological futures. The most carbon-intensive scenario foresees population and economic growth driving atmospheric carbon dioxide levels to 1370 parts per million (pmm) by 2100. That's more than three times the current level of 380. The rosiest scenario has them cresting at 490 ppm before declining.

Starting with only four pathways should actually generate better data, according to researchers. With fewer total emissions scenarios to model, scientists will be able to do more runs on the same scenarios. That allows so-called ensembles in which models are repeatedly run with nearly identical starting conditions, yielding trends and reducing uncertainty. Delegates in Budapest also endorsed running climate models out to 2035 in addition to continuing the century-long projections. The computing power saved by modeling shorter time periods will allow scientists to get results on smaller geographic grids, giving more local details. "The focus will be on regional climate change and extremes," says NCAR's Gerald Meehl.

The new process is expected to foster greater collaboration, says Hibbard, describing the previous regime as one of "toss your data over the fence and see ya later." For example, climate modelers would like to know how reforestation might affect Earth's albedo this century. But the success of future reforestation efforts depends on technological and policy changes forecast by economic modelers. Similarly, climate modeling could forecast damage by drought to agriculture, which scenario builders could use to later revise the scenarios that underpin the whole enterprise.

The panel will meet again in September in Geneva to choose its leaders—current IPCC chair Rajendra Pachauri wants to stay on and enjoys widespread support—and the heads of the three working groups. In the meantime, scientists seem to be happy with their new marching orders. "This whole thing is turning on new steam," says Hibbard, who has already e-mailed details of the IPCC's decisions to a modeling team at NCAR.

–ELI KINTISCH

FOCUS

Carbon-friendly
synfuels?

306

Mapping the
flu virus

310

GLACIOLOGY

Greenland Ice Slipping Away but Not All That Quickly

Almost 6 years ago, a paper in *Science* warned of an unheralded environmental peril. Melted snow and ice seemed to be reaching the base of the great Greenland ice sheet, lubricating it and accelerating the sheet's slide toward oblivion in the sea, where it was raising sea level worldwide (12 July 2002, p. 218).

Now a two-pronged study—both broader and more focused than the one that sounded the alarm—has confirmed that meltwater reaches the ice sheet's base and does indeed speed the ice's seaward flow. The good news is that the process is more leisurely than many climate scientists had feared. "Is it, 'Run for the hills, the ice sheet is falling in the ocean'?" asks glaciologist Richard Alley of Pennsylvania State University in State College. "No. It matters, but it's not huge." The finding should ease concerns that Greenland ice could raise sea level a disastrous meter or more by the end of the century. Experts remain concerned, however, because meltwater doesn't explain why Greenland's rivers of ice have recently surged forward (*Science*, 24 March 2006, p. 1698).

The original paper described work by glaciologist Jay Zwally of NASA's Goddard Space Flight Center (GSFC) in Greenbelt, Maryland, and colleagues. They noted that whenever summer warmth swelled, producing more meltwater on the ice sheet, the ice beneath Swiss Camp—70 kilometers inland of the west coast—sped up as much as 28% on its 115-meter-per-year creep toward the Davis Strait. Presumably, the meltwater disappearing down tubular chasms called moulins somehow reached the ice sheet's base a kilometer down and slicked it up, letting the ice take off, if only for a couple of months.

But no one was sure how meltwater managed the feat or whether the ice acceleration at Swiss Camp continued all the way to the sea. So glaciologists Ian Joughin of the University of Washington's Applied Physics Laboratory

in Seattle, Sarah Das of Woods Hole Oceanographic Institution in Massachusetts, and their colleagues took a dual approach to meltwater lubrication, as they report in two papers published online this week in *Science* (www.sciencemag.org/cgi/rapidpdf/1153288.pdf). They took a close look at how lubrication works by instrumenting a growing puddle of meltwater south of Swiss Camp. For the broad view, they went to images made from satellite-borne radar that tracked ice sheet movement every 24 days across a 425-kilometer-by-100-kilometer swath of the west coast, including Swiss Camp.

The meltwater monitoring caught a 4-kilometer-long, 8-meter-deep lake disappearing into the ice in an hour and a half. As theorists had supposed, once the lake water was deep enough, its weight began to wedge open existing cracks, which only increased the weight of overlying water on the crack tip and accelerated cracking downward. Once the main crack reached the bottom of the ice, heat from churning water flow melted out parts of the fracture, and drainage took off. The lake disappeared in about 1.4 hours at an average rate of 8700 cubic meters per second, exceeding the average flow over Niagara Falls. That's almost four Olympic pools a second.

Down the hatch. Meltwater pouring into moulins like this one can lubricate the ice sheet's base, but the ice's resulting acceleration into the sea is modest.

For all the lake's water dumped under the ice that day and all the water drained into new moulins in the following weeks, the ice sheet moved only an extra half-meter near the drained lake. Das and her colleagues conclude that an unknown number of lake drainages and new moulins account for the acceleration reported by Zwally and now confirmed more broadly by the radar observations. Joughin and his colleagues report that in August 2006 the ice sheet sped up over a broad area by a hefty 48% above its 76-meters-per-year mean speed as melt lakes grew—and disappeared—under summer warmth.

The good news came toward the coast, where the ice speeds up as the flow narrows into a few outlet glaciers that deliver the ice to the sea. Those glaciers moved only 9% faster than normal in August of 2006. "Meltwater does indeed cause substantial speedup" inland on the ice sheet, says Joughin, "but it has a small effect on outlet glaciers." That may be because the beds on which outlet glaciers slide are already smooth and well lubricated year-round, the group speculates. All in all, meltwater lubrication "likely will have a substantive but not catastrophic effect on the Greenland Ice Sheet's future evolution," the group reports.

Alley agrees. "Could things go two times faster [due to meltwater] than we thought 10 years ago?" he asks. "Yes. They can go faster but not ridiculously faster." The danger now, warns glaciologist Robert Bindschadler of GSFC, is "falling into the same 'We now know how ice sheets work' trap that my generation was guilty of 5 years ago." After all, some of Greenland's outlet glaciers began galloping to the sea in recent years. If not meltwater, what set them off?

–RICHARD A. KERR

ARCHAEOLOGY

Team Unveils Mideast Archaeology Peace Plan

Last week in Jerusalem, a small team of Americans, Israelis, and Palestinians presented a peace plan for the Holy Land's archaeological riches. After secretly meeting in different countries over the past 5 years, the eight archaeologists offered their view on the fate of thousands of artifacts and sacred sites. Their aim is to remove the divisive issue from negotiations over Palestine's future as an independent state.

But whether their hard work will pay off is anyone's guess. "I am doubtful that an unofficial document drawn up by some well-meaning archaeologists will make any difference," says Patrick Daly, an archaeologist at the Asia Research Institute in Singapore who has been visiting faculty at An-Najah National University in Nablus in the West Bank.

At the heart of the controversy is the question of what should be done with material

Waiting for peace. The Temple Mount will be at the epicenter of a protective "Heritage Zone" in Jerusalem.

removed from the Palestinian West Bank territory since the 1967 Arab-Israeli War. During decades of settlement building, the Israeli Antiquities Authority (IAA) has uncovered and removed artifacts including coins from the Crusades and stone tools from the Paleolithic. When—and if—a Palestinian state is created, the question will become whether some or all of those objects, most now stored in Israeli museums and warehouses, will be repatriated. Another contentious issue is who will maintain and control access to religious sites on either side of the border, particularly in and around Jerusalem, a city claimed by both sides. These issues are a "major hurdle for peace," says Ran Boytner, an Israeli-born archaeologist at the University of California, Los Angeles (UCLA).

So starting 5 years ago, Boytner and Lynn Swartz Dodd, an archaeologist at the University of Southern California (USC) in Los Angeles, quietly assembled a team of Israeli and Palestinian colleagues. "People who participated did so at great risk," says Boytner. Israeli academics collaborating with Palestinians, and vice versa, are often viewed as traitors, he says, and losing one's job—or life—is a real possibility. Most of the team made their names public last week. But one of the three Israelis and one of the three Palestinians remain anonymous.

The meetings were initially held in Vienna—"neutral" ground, says Dodd—then in Southampton, U.K., and finally in Jerusalem. Expenses were covered by a $150,000 fund created by USC, UCLA, the Washington, D.C.–based United States Institute for Peace, and other donors. Despite often tense negotiations—professional facilitators were brought in several times to keep the process going—the team grew intimate. "Our meetings usually included at least one recent-pictures-of-kids swap," Dodd says. "It's probably no accident that all participants are parents who are thinking toward their children's futures."

The first challenge was to account for "tens of thousands of artifacts" and nearly 6000 sites, says David Ilan, one of the Israeli participants who directs the Nelson Glueck School of Biblical Archaeology in Jerusalem. "Given construction activities in the West Bank, including the building of the separation fence, new [archaeological] sites are discovered there almost daily," says team member Raphael Greenberg of Tel Aviv University in Israel.

Records from West Bank excavations were hard to come by. After being rebuffed by IAA, Greenberg sued and won a court injunction to obtain the data. "This filled in many gaps," says Ilan. Boytner says all the data "will soon be made available to the public." (IAA declined to comment.)

The team's plan calls for a protective "Heritage Zone" around the oldest part of Jerusalem, extending to the city's 10th century boundaries. Archaeological sites in the zone would be accessible to anyone, and any research would have to be done with full transparency. The plan also recommends the repatriation of all artifacts found since 1967 to the state in which they were unearthed—essentially a one-way transfer from Israel to Palestine. To house all the material returned to the Palestinian side, new museums and conservation laboratories would be created. Exactly who would construct the facilities is not spelled out, but Katharina Galor, an archaeologist at Brown University who is "not very optimistic" for the plan's future, estimates the cost at "millions if not billions of dollars."

About 50 Israeli archaeologists, including IAA officials, showed up on 8 April in Jerusalem to hear the U.S. and Israeli part of the team make their case, says Boytner. (No Palestinians attended.) Ilan was prepared for the worst but says "surprisingly, the overwhelming response was positive and congratulatory. Not a single person spoke against the document." The consensus was that "this process should continue," says audience member Hanan Eshel of Bar-Ilan University in Ramat-Gan, Israel. The buzz at the meeting was that the team's anonymous Israeli member is an IAA archaeologist. "We will not comment," says Boytner.

A follow-up meeting is being planned for the Israeli side. Among Palestinians, there is broad support but also those who "do not want to involve Israel whatsoever in a future Palestinian state," says team member Ghattas Sayej, an archaeologist with the Palestinian Association for Cultural Exchange in Ramallah, the West Bank. The effort to convince Palestinian archaeologists to formally ratify the plan is being led by team member Nazmi el-Jubeh, co-director of RIWAQ, an architectural conservation organization in Ramallah.

And after that? "It's up to our politicians," says Sayej. "The plan is there."

–JOHN BOHANNON

Swamped. Potential cuts in U.S. support for research may slow the introduction of new rice varieties such as those tolerant of submergence.

INTERNATIONAL AID

As Food Prices Rise, U.S. Support For Agricultural Centers Wilts

A recent spike in wholesale and market prices for rice, wheat, and maize has touched off food riots and prompted countries with surpluses to impose restrictions on grain exports. Food importers are in a panic, and relief organizations are warning of a pending calamity. In response, U.S. President George W. Bush earlier this week ordered up $200 million in emergency food aid. Behind the scenes, however, researchers charge that the U.S. government is moving to slash funding for international agricultural research.

"You couldn't ask for worse timing," says Robert Zeigler, director general of the International Rice Research Institute (IRRI) in Los Baños, Philippines. "Part of the reason we're having this deterioration of the global agricultural situation is that there has been a steady erosion of support for research." IRRI has put a freeze on hiring and is holding back on planned research investments until the budget is confirmed.

Last week, several concerned scientists circulated an online petition seeking to reverse cuts to research funds they say are being planned by the U.S. Agency for International Development (USAID), calling them "unacceptable mistakes that will damage worldwide food production for many years to come." The group argues that international agricultural research should be expanding. "Restoring [support] isn't really enough; this should be an area of major growth," says Jeffrey Bennetzen, a plant geneticist at the University of Georgia, Athens, and a petition organizer.

In 2006, USAID provided about $56 million to a network of 15 centers around the world called the Consultative Group on International Agricultural Research, or about 12% of CGIAR's budget. According to CGIAR spokesperson Fionna Douglas, USAID officials warned in February of a strong likelihood of "a huge cut" in funding this year. CGIAR later learned it could be as much as 75%; the final figure is still being determined. Douglas says they've been told that the 2008 USAID budget includes extensive earmarks requiring funding to be directed primarily to health issues, leaving little for agriculture. USAID officials did not respond to requests for details on CGIAR funding charges.

In the meantime, a perfect storm is brewing. Across the developing world, farmland and water for irrigation have been lost to urban development and industrialization. Grains are being diverted to feed livestock to meet rising demand for meat and to make biofuels. Droughts in Asia and Australia have severely curtailed grain production. And productivity has stagnated, says Zeigler, due to cuts in agricultural research in the 1990s.

The result is a steady rise in grain prices. On 20 March, the U.N.'s World Food Programme issued an appeal for help in covering a $500 million shortfall in its $2.9 billion budget this year to feed 73 million people in 78 countries. In the 3 weeks since, food prices shot up another 20%. "You could see the train wreck coming for years," says Zeigler.

Bennetzen and his colleagues plan to send their petition (www.ipetitions.com/petition/cgair_support), which has garnered more than 600 signatures, to key members of the U.S. Congress and USAID administrators. With recent headlines drawing attention to the food crisis, says Bennetzen, "I don't think it should be difficult to communicate that it's a desperate situation."

–DENNIS NORMILE

Come Get Your Stem Cells

BERLIN—Scientists in Germany will soon have more human embryonic stem (ES) cell lines available for their research. On 11 April, lawmakers voted 346 to 228 to allow them to work with cells derived before 1 May 2007. Germany's embryo protection law makes it illegal to derive human ES cells, and previously German scientists were only allowed to work with imported human ES cells that had been derived before 1 January 2002. The new law means that more than 500 cell lines are now legal for import instead of just 21. It also clarifies that German researchers working abroad will not be prosecuted for working with cell lines that are illegal in Germany. Developmental biologist Hans Schöler of the Max Planck Institute for Molecular Biomedicine in Münster says the wider choice is important because different ES cell lines have different properties. It will also make international collaborations much easier, he says.

–GRETCHEN VOGEL

Institute Called a Troubled Environment

A new National Institutes of Health (NIH) report identifies several problems with the way the National Institute of Environmental Health Sciences (NIEHS) functioned before Director David Schwartz resigned in February. Last year, a Senate investigation raised conflict-of-interest issues about his consulting for law firms, overspending his personal lab budget, and collaborating with Duke University, his former employer. Schwartz agreed to step aside in August during a management review requested by the House Appropriations Committee. In a 47-page report dated 9 April, the NIH Office of Management Assessment found that NIEHS officials failed to document why they funded 45 grants over 2 years that were not among the 2500 top-rated proposals. It also says that an understaffed ethics office was unable to read most conflict-of-interest reports filed by NIEHS scientists. In a rare reference to Schwartz, the report says that staffers managing his conflicts were put in "a very difficult" position because they reported to him.

Senator Chuck Grassley (R–IA) complained in a letter to NIH that the report ignores the fact that NIH Deputy Director Raynard Kington, not the NIEHS ethics officials, approved Schwartz's outside activities. Kington says the suggestion that he broke rules is "preposterous." NIH says other reviews of Schwartz's conduct are ongoing.

–JOCELYN KAISER

SCIENTIFIC PUBLISHING

Croatian Editors Fight With Medical School Over Journal's Fate

At an international meeting on research integrity last fall, Ana Marušić spoke on the problems facing small journals. Since then, the anatomy professor has many more stories to tell. She and her husband, Matko Marušić, co-editors-in-chief of the *Croatian Medical Journal* (*CMJ*), are at the center of a controversy that threatens their jobs, observers say, the journal's independence and its example of quality scientific publishing in countries outside the scientific mainstream.

The editors—both professors at the University of Zagreb Medical School, which, along with three other Croatian medical schools owns the journal—have faced charges of plagiarism and defaming the university. The Marušićs say they are being targeted for their insistence on research ethics and for bringing to light corruption and plagiarism in the Croatian medical community. Their critics, primarily academics at the same school, charge that it's the editors who are behaving unethically and destroying the trust essential for the journal's operation.

Founded in 1991, *CMJ* was conceived as a forum for doctors ensnared in the country's civil war to communicate to the outside world. Non-Croatian authors were also welcome. From the beginning, says Ana Marušić, the goal was to educate scientists from devel-

oping countries on how to communicate their work better—primarily in English. Available for free online, the journal is listed in major citation indexes. Ana Marušić is currently president of the Council of Science Editors,

Under fire. *Croatian Medical Journal* editors-in-chief Matko and Ana Marušić worry about their journal's independence and their careers.

an international organization of journal editors, and also headed the World Association of Medical Editors. "Considering the size of the country and the resources available, they really hit the ground running," says Mary Scheetz, an expert in research ethics and scientific publishing at the University of Virginia, Charlottesville. "Their journal became respected."

That respect has perhaps contributed to

the editors' problems. In Croatia, professors must publish at least five papers in journals indexed in *Current Contents* to receive promotions. Because *CMJ* is the only Croatian journal listed in *Current Contents*, a rejection can thwart careers. In part because of that, says Davor Solter, director emeritus at the Max Planck Institute of Immunobiology in Freiburg, Germany, and member of *CMJ*'s advisory board, the Marušićs "have a lot of friends outside Croatia and a lot of enemies inside Croatia."

The Marušićs say their troubles started in 2001 when the journal rejected a paper by a Zagreb colleague, based on unfavorable reviews. They claim their problems worsened after a September 2006 commentary in the *British Medical Journal* (*BMJ*) detailed two examples of plagiarism by Asim Kurjak, a prominent gynecologist at the University of Zagreb. Ana Marušić knew the commentary's author, Iain Chalmers, editor of the James Lind Library in Oxford, U.K., through publishing circles, and she says that Kurjak and others accused the couple of prompting the *BMJ* article. Chalmers says a chance conversation with a Norwegian researcher prompted him. He says he checked some of the facts with another Croatian scientist but deliberately avoided mentioning it to the Marušićs.

After pressure from the Croatian press, Nada Čikeš, dean of the Zagreb medical school, referred the Kurjak matter to the school's Court of Honor. In October, it found that Kurjak had behaved unethically, but because he had retired a month earlier, the

REGULATORY SCIENCE

Changes to EPA Toxicology—Speed or Delay?

The U.S. Environmental Protection Agency (EPA) has substantially modified the way it updates a database on chemical hazards that influences how chemicals are regulated. The agency says the changes should make the process more transparent and more rigorous, and speedier. But critics argue that the new procedure is more secretive and gives too much clout to federal agencies that pollute or face massive cleanup costs. One result, they say, will be further delays in regulation.

Begun in 1985, the Integrated Risk Information System (IRIS) contains EPA scientists' appraisals of the chronic health effects of more than 540 chemicals. EPA

regulators use this information to revise drinking water standards, for example, or set cleanup levels at Superfund sites. Many states and agencies around the world also use the data. "IRIS is the gold standard," says Jennifer Sass of the Natural Resources Defense Council (NRDC) in Washington, D.C.

IRIS has sometimes been a battleground. For example, the Department of Defense (DOD) criticized the science behind EPA's draft IRIS document for a solvent called trichloroethylene, which contaminates many military bases. This 2001 draft identified stronger evidence of car-

cinogenicity and could ultimately force DOD to spend billions more cleaning up polluted aquifers. Since then, the White House's Office of Management and Budget (OMB) started reviewing all IRIS drafts. Many came back with edits that made health effects seem more uncertain, says an EPA official who asked to remain anonymous, whereas other drafts remain in limbo at OMB. On average, documents now take 5.5 years to complete, and the most controversial can stretch to more than a decade.

Three years ago, EPA Administrator Stephen Johnson asked his Office of Research and Development (ORD), which runs IRIS, to make the process more predictable and transparent. Under changes announced on 10 April, the public and

Impact factor. Started in war, the *CMJ* has admirers around the world

court did not punish him. About the same time, Čikeš also asked the body to look into plagiarism charges involving Ana Marušić.

In May 2006, an anonymous letter to Croatian science ministry and university officials said that significant portions of a 2002 anatomy textbook co-authored by Ana Marušić were identical to passages in an American textbook. This month, the Court of Honor issued a "public warning" to Ana Marušić. It found her "responsible for the fact that the textbook … without providing the source and authors, copied and translated English text."

Ana Marušić does not dispute the lack of acknowledgement but says she relied on her publisher to seek the necessary permission; the publisher said it tried several times before deciding the American publishers did not object. "It was definitely a mistake" not to credit the source textbook, she says.

Matko Marušić has also run afoul of university authorities. Even his supporters say the editor is an intense person whose energy and stubbornness can rankle. Last week, a disciplinary hearing was held to decide whether comments he made to a newspaper about corruption in the Croatian scientific community defamed the university. As part of that process, the committee asked three university psychiatrists for their opinion of Matko Marušić's public

comments and his correspondence. One of them told *Science* that he declined to cooperate, but Matko wrote to the American Psychiatric Association protesting the school's request.

CMJ's owners, the deans of Croatia's four medical schools, are now considering a proposal by Čikeš to put the journal under their direct control instead of the current eight-member management board. Čikeš has also proposed that *CMJ*'s editor-in-chief be rehired and that anyone who had been reprimanded by his or her university should be disqualified. Čikeš says the changes would bring the journal in line with governance standards recommended by the World Association of Medical Editors. The Marušićs have never been formally evaluated or elected to their positions, she says. She does acknowledge their accomplishments. "I am happy and proud that we have such a good journal," she says. However, she says, the ongoing disputes have gotten out of hand. "The whole thing is immobilizing parts of the institution."

Solter is dismayed by the fight. Given the relative success of the journal, the disputes seem like a waste of time and energy, Solter says: "When all is said and done, they made the journal what it is. … To get that done, maybe you have to be a bit obnoxious. The journal and its editors should be left alone to do their work."

–GRETCHEN VOGEL

interested federal agencies now have a chance to comment on IRIS's early "qualitative" drafts. In an added step, federal agencies—but not the public—will get a confidential "sneak preview" of the final draft before it is sent to peer review.

DOD says it's pleased with the changes. The early reviews will enable DOD to resolve questions of scientific uncertainty more quickly and get a head start on managing risks, such as by finding substitute chemicals, predicts Shannon Cunniff, who directs DOD's program on emerging contaminants.

NRDC's Sass worries that the added review will let federal agencies delay the process. She's also concerned that interagency comments on drafts that were previously part of the public record will now be secret. Cunniff says, however, that DOD plans to make public the scientific feedback it sends to EPA. And George Gray, who heads ORD, emphasizes that the final decisions on the content of IRIS documents will remain in EPA's hands. "Anything we do has to be scientifically justified," he says.

Skeptics remain. In a statement, Senator Barbara Boxer (D–CA) called the changes "devastating" and announced that the Environment and Public Works Committee, which she chairs, plans to conduct an oversight hearing on EPA's toxics program. In addition, the Government Accountability Office will shortly release a study she requested on political influence on IRIS.

–ERIK STOKSTAD

Obama Questions NASA Programs

Democratic presidential candidate Senator Barack Obama (D–IL) says that NASA "needs to be redefined." That may be cause for space scientists to cheer. Answering a question 11 April from a high school student in Columbus, Indiana, Obama said he was "a big supporter of the space program" but that as president he would ask whether the agency should emphasize robotic probes instead of human launches. Such probes, he noted, "oftentimes are cheaper and less dangerous but yield more information." Obama says he wants a "major debate" on the subject with implications for the agency's budget. Earlier this year, Obama backed construction of a new human launcher, after previously calling for a 5-year delay.

Meanwhile, NASA decided this week to extend the Cassini mission in the Saturn system by 2 years. The agency planned to shut down the spacecraft in July, but an impressive flow of data on the planet's moons persuaded managers to keep it operating until 2010.

–ANDREW LAWLER

Whole-Genome Studies: You Are Wellcome

British researchers are expanding their search for genes involved in common diseases. Over the next 2 years, the Wellcome Trust Case Control Consortium (WTCCC) will genotype 120,000 people and probe the data for genetic markers involved in 25 disorders. Last year, the consortium provided some of the flood of so-called genome-wide association studies finding genetic markers involved in disorders such as heart disease or diabetes. The consortium's new £30 million effort, funded by its charity namesake, will now study the DNA of seven times as many people in the United Kingdom and other countries.

–JOCELYN KAISER

Cleaner Ships Ahoy

In an effort to improve air quality in ports and coastal areas, the International Maritime Organization (IMO) plans to cut sulfur pollution from ships. Earlier this month, IMO lowered the current standard from 4.5% sulfur content in fuel to 3.5% by 2012. Most ships already use fuel with only 2.5% sulfur, but the standard becomes 0.5% by 2020. That's still more than 10 times higher than European Union diesel right now, says Jackie Savitz of the advocacy group Oceana. "The U.S., or anyone really, can't afford to rely on IMO to address these types of problems." **–ERIK STOKSTAD**
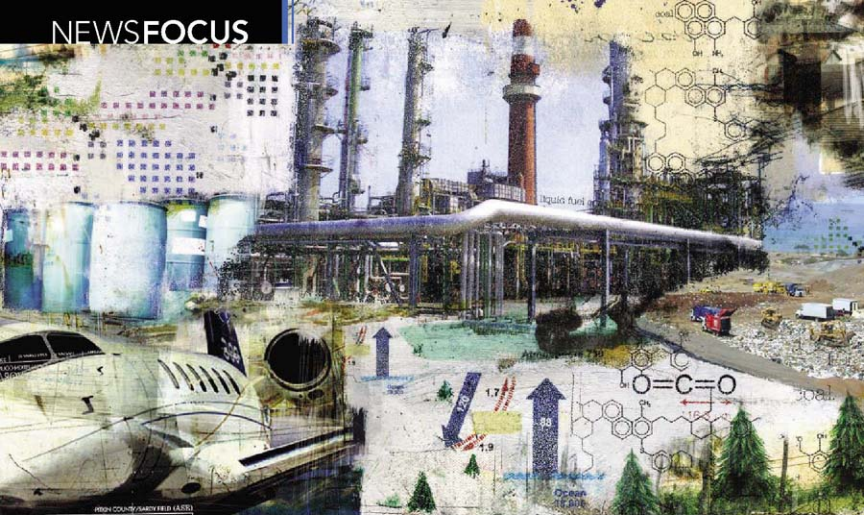
The Greening of Synfuels

An old, dirty technology to make transportation fuels from coal could fight global warming, say proponents. The trick is using more biomass and burying the carbon dioxide that's generated

A multibillion-dollar U.S. effort to turn coal into gasoline was a colossal flop in the 1980s, plagued by mismanagement, political wrangling, and falling oil prices. Environmentalists concerned about the impact of additional coal mining cheered the end of the synthetic fuels program, which was aimed at cutting U.S. dependence on oil from the Middle East.

A generation later, the geopolitical reasons for reducing U.S. oil imports are more compelling than ever. And with oil prices above $100 a barrel, the economic equation has changed. So it's no surprise that a few U.S. energy companies have drawn up plans for synfuels plants that would produce millions of barrels of the alternative fuel annually.

But this time around, the technology is also gaining support from a seemingly unlikely source. A group of climate scientists believes that, barrel for barrel, synfuels can emit less carbon dioxide (CO_2) than oil and, at some point, even reduce the amount of carbon in the atmosphere. "When you make synfuels, you have an incredible opportunity" to tackle climate change, says Princeton University physicist Robert Williams, an advocate of the technology.

Living up to that promise won't be easy, however. The two keys to making synfuels green are using large amounts of plant biomass along with coal and storing in the ground the CO_2 emitted during the production of synfuels. And neither has been implemented on a commercial scale. Most environmental groups are still horrified by the thought of more synfuels plants and are loath to see coal mining expanded. They also point out that the process produces CO_2 at twice the rate of making gasoline from crude oil without CO_2 storage and the use of biomass, the result would be disastrous. At least eight synfuels plants are expected to open soon in China, with 17 more planned; they will spew forth millions of tons of CO_2. A coal-fed facility in Secunda, South Africa, built to cope with an apartheid-era fuel embargo, is the planet's single biggest point source of carbon, emitting 20 million tons of CO_2 a year.

"[Synfuels] may be worth looking into, and I have no doubt someone's going to make money with the process," says energy professor Daniel Kammen of the University of California, Berkeley. But he thinks those who see a climate benefit are underestimating the costs of large-scale carbon storage while overestimating the availability of biomass that can be harvested without having deleterious effects. As a climate solution, he says, "I'm a lot less sanguine that it's going to work out."

What a gas

The chemistry involved in making synfuels is not complicated. The process begins by turning coal into gas, which creates carbon monoxide and hydrogen (see diagram, p. 309). The resulting syngas, as it's called, is then converted with catalysts into products such as diesel fuel, jet fuel, or chemical feedstocks after scrubbing for pollutants.

Germany operated the first large-scale commercial synfuels plants in the 1940s to provide fuel for a Nazi war machine starved by an Allied oil embargo. Then the 1970s gasoline crunch led U.S. President Jimmy

Carter and Congress to create the $20 billion Synthetic Fuels Corp. in 1980. The goal was to use coal to produce 700 million barrels of oil per year by 1992. The corporation spent $2 billion on demonstration projects in California, Louisiana, and North Dakota. But management scandals, battles between the corporation and the White House during the Reagan Administration, and, ultimately, the falling price of oil—it hit $21 a barrel in 1986—caused Congress to pull the plug that year. Experts said the only thing that would revive synfuels was $100-a-barrel oil.

And here we are. In the United States, two companies lead the Synfuels 2.0 effort. Baard Energy, based in Vancouver, Washington, hopes next year to begin building a $5 billion plant in Wellsville, Ohio, that would produce 50,000 barrels a day of diesel, jet fuel, and other chemicals. Rentech Inc., based in Los Angeles, California, hopes to open a plant in Natchez, Mississippi, in 2011 that would eventually make 30,000 barrels of fuel a day. Although the hefty price tag of a synfuels plant makes it less likely that enough will be built to have a major impact on global transportation needs, Baard's owner and founder, John Baardson, says the plant will make money as long as the cost of a barrel of oil remains above $50.

The companies plan to use 30% and 10% biomass by weight, respectively, and store the CO_2 they make underground. That mix, they say, will produce fuels with a life cycle carbon footprint much smaller than the one left by those derived from Middle Eastern oil. Future projects using greater proportions of biomass, advanced gasifiers, and carbon storage could result in a carbon-negative process, say proponents, storing indefinitely the CO_2 that plants had taken up from the atmosphere. Baard says that getting enough biomass for its Ohio plant won't be a problem. Rentech hopes to be able to use garbage, which is also plentiful. (A third

company, owned by the power utility DKRW, is planning a coal-only gasification project in Wyoming that will inject CO_2 as well.)

A Dutch utility called Nuon has been pioneering this method, gasifying an 80-20 mix of coal and wood chips since 2006. (Its plant in Buggenum, Netherlands, generates power instead of fuel, but the gasification step is identical.) "They've solved a number of technical problems," says Baardson, including selecting the best feedstocks and preparing them for conversion.

Unlike coal, which is easily ground into tiny spheres, the fibrous wood gets stuck as it is fed into the gasifier, creating an uneven flow. Dutch engineers have developed a way of mixing the two feedstocks to make them flow better. A new process of drying and charring the wood beforehand, developed by the Energy Research Centre of the Netherlands, has also helped keep the mixture flowing evenly into the gasifier. The process requires extra energy, but by reducing the weight of the material it lowers transportation costs.

Engineers in industry believe that preparing the biomass is the main technical hurdle to gasifying it and cite Nuon's success as proof. But with Nuon keeping its methods secret, government researchers want to explore the new feedstock more. "We don't exactly know how biomass is going to affect the gasifier, gas cleanup, or catalysts systems," says Daniel Cicero of the National Energy Technology Laboratory in Morgantown, West Virginia. The lab announced a $7 million research program last month to identify the minerals found in biomass feedstocks such as poplar or switchgrass and to examine how they might affect the system.

Nuon's gasifier, built by Shell, operates

> *"The thing that makes [synfuels] so bad for the climate could make [them] so good for the climate."*
>
> —DANIEL SCHRAG, HARVARD UNIVERSITY

above 1200°C. That temperature melts the inorganic ash that the process creates. But gasifiers that run at temperatures hundreds of degrees cooler could save in construction and operating costs, says Richard Bain of the National Renewable Energy Laboratory in Golden, Colorado.

Cooler gasifiers have their own problems, however. Lower temperatures mean that less of the feedstock—be it coal or biomass—is converted into syngas. The toxic, carbonaceous muck that remains is costly to dispose of. Researchers hope that better computer modeling and new chemical techniques will help them more fully process the gunk.

New chemistry could cut costs even more dramatically. Two years ago, chemical engineer Lanny Schmidt of the University of Minnesota, Minneapolis, demonstrated how to gasify biomass by releasing tiny bits onto a catalyst made of rhodium and cerium, whereby it is converted instantly to syngas in an oxygen-rich vessel (*Science*, 3 November 2006, p. 801). Its industrial advantages include shortening the duration of the process—to roughly a tenth of the time of existing gasifier designs—and leaving behind almost no carbon. Because the reaction continually releases its own heat—700°C—the technique could eliminate costly external heating. But Schmidt acknowledges he still needs to solve a copious "ash problem" before synfuels plants can be shrunk to the size to ethanol facilities, which are small enough to sit adjacent to local farms.

Going under

To store the CO_2 that synfuels plants create, researchers hope to take advantage of the fact that the process creates a concentrated CO_2

It's a gas. Traditional synfuels plants take coal and turn it into syngas. The gas is then catalyzed into various liquid fuels. Proposed plants would also store underground the CO_2 that is created. Greater reliance on biomass would make the process more carbon friendly.

Gasifier

Catalyst

Syngas

$H_2 + CO$

CO_2

Coal

Waste/biomass

Diesel fuel

Jet fuel

Underground sequestration for oil recovery
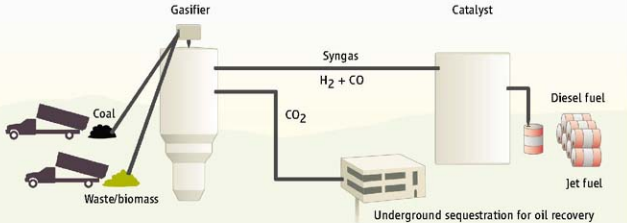
A burning question. South Africa's Secunda facility is the world's biggest point source for carbon emissions, but synfuels can be made cleaner.

A Fluctuating Carbon Footprint

GRAMS CO₂ PER KM DRIVEN

-400 -200 0 200 400 600

Gasoline

Coal synfuel, no CO₂ storage

Coal synfuel, CO₂ storage

Biomass synfuel, no CO₂ storage

Biomass synfuel, CO₂ storage

stream that can simply be injected into deep underground formations. In contrast, CO_2 from a standard generating plant must be separated from other flue gases (*Science*, 13 July 2007, p. 184). "In a way, the thing that makes [synfuels] so bad for the climate could make [them] so good for the climate," says Daniel Schrag, a Harvard University geochemist who works as a part-time consultant for Rentech. Capturing and storing a ton of carbon from a standard coal plant would cost $40, according to a survey last year by a team of researchers at the Massachusetts Institute for Technology in Cambridge. Rentech says its Mississippi plant, strategically located near pipelines that currently bring CO_2 to oil fields, will do it for a net of $6 a ton.

But the amount of CO_2 needed to be stored by a new generation of synfuels plants dwarfs current experimental efforts. The three largest projects worldwide—in Algeria, off the coast of Norway, and in Saskatchewan, Canada—are each storing roughly 1 million tons per year. Baard expects its plant alone will produce more than four times that amount. Baard and Rentech plan to sell the CO_2 from their plants to oil companies to help them squeeze the last drops out of existing wells, a process that geologists say effectively stores the CO_2 once the wells are sealed. But such opportunities are relatively rare. Fortunately, there's plenty of available space elsewhere: A government survey last year found that the United States has room underground and near power plants for at least 91 billion metric tons of CO_2, enough to absorb many decades of emissions.

The process involves injecting a stream of CO_2, liquefied by high pressure, into a series of wells drilled thousands of meters into porous rocklike sandstone. The formations are capped with impermeable layers of rock. Inside the space, the liquid CO_2 displaces briny liquids as it fills pores. Results of early tests on a small scale have been positive, but scientists say they still have a lot to learn as they scale up injections. "I wouldn't say there are any major technical [barriers]," says engineer Sean McCoy of Carnegie Mellon University in Pittsburgh, Pennsylvania. "We want to make sure there aren't any surprises."

To reassure the public that underground carbon sequestration is reliable and safe, hydrologist Diana Bacon of Pacific Northwest National Laboratories in Richland, Washington, says researchers need better computer models of how stored CO_2 behaves. Adding complex geochemistry to the models is a first step. Liquid CO_2 under pressure, for example, can cause the formation of solid salts such as sodium chloride that can block pores and alter the flow of the injected CO_2. Varying mixtures of calcium, dolomite, and sandstone found in deep sedimentary rocks could affect CO_2 behavior differently, says Bacon. Injecting CO_2 into the basalts found beneath much of the United States and India, among other places, can have similarly hard-to-model effects.

Megascale synfuels projects would give engineers the experience they now lack in long-term sequestration of CO_2. "We need to just get moving," says Bacon. But that's hard to do in the United States, where pure CO_2 streams are relatively rare despite the heavy use of fossil fuels. China's projected synfuels plants give that country the chance to become "a world leader" in CO_2 storage, says

Princeton's Williams. But despite nascent partnerships with the U.S., European Union, and U.K. governments, the only large-scale test announced so far in China is a $1 billion power plant, dubbed GreenGen, in Tianjin. Several government-owned companies expect to begin construction next year.

Companies say that synfuels could become an important energy source sooner if the U.S. government lends a hand. One potentially huge customer for synfuels is the U.S. Air Force, whose planes now consume 11.4 billion liters of fuel a year. Synfuels makers want Congress to grant the Pentagon the authority to sign long-term fuel-purchasing contracts for synfuels. The lawmakers who oversee the Pentagon have been mum on the matter. A compromise requiring restrictions on carbon emissions for federally supported synfuels seems possible, although a similar deal involving tax breaks and production credits failed last year.

Environmental groups oppose such a deal. David Hawkins of the Natural Resources Defense Council in Washington, D.C., fears that any legislation will open the door to a surge in synfuels made purely from coal. Even if the CO_2 generated could be stored, he says, the effects of expanding coal mining could be extremely harmful to the environment.

Notwithstanding the technical hurdles, the fate of synfuels may hang on whether companies are forced to pay a price for the carbon they emit into the atmosphere. Opponents point out that Rentech and Baard, notwithstanding their pledges, are free to use only coal in their synfuels plants and emit millions of tons of CO_2 per year. "That's why a problem if you don't have a price on carbon," Schrag counters. Since they can use nonfood crops or plant waste, synfuels "can be better than ethanol," he adds, citing the negative impact of corn ethanol on food prices and its projected deleterious effect on climate (*Science*, 29 February, p. 1235). Economic policies that reward good behavior will not only serve as a huge incentive to the synfuels industry, he notes, but also have more global effects: "If there isn't a carbon price, we're not going to solve the climate problem anyway."

–ELI KINTISCH
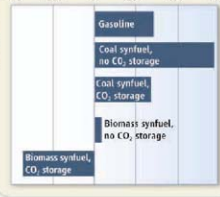
OCEANOGRAPHY

Watery Echoes Give Clues to the Past and Future of the Seas

A handful of oceanographers and geophysicists are recording seismic whispers of the ocean's structure

Last spring, Katy Sheen listened to the sounds of the ocean from a ship off the coast of Spain. A relaxing vacation? Hardly. Sheen, a graduate student at the University of Cambridge in the U.K., is one of a handful of scientists adapting a technique called seismic profiling to oceanography. By observing the changing speeds of sound waves propagating through water, geophysicists and oceanographers hope to extract information about the ocean's temperature, salinity, and velocity.

Geologists and oil companies have long used ship-based seismic profiling to probe density changes in the solid earth beneath the sea, but the technique of mapping the ocean's internal structure this way is less than a decade old. If efforts like Sheen's succeed as expected, scientists will gain a powerful new tool that could unlock the volume of the ocean to rapid and remote study, much as satellites did for the ocean's surface. "When satellite observations came along, the oceanographic community … said, 'Well, it's not going to tell us anything new,' but it did and it was important," says Nicky White of the University of Cambridge, Sheen's supervisor.

Particularly exciting, White and others say, is the prospect of tapping a jackpot of knowledge from decades of "legacy data" that energy companies have gathered while sounding sea floors in search of oil. The results could prove invaluable for measuring how the ocean's waters interact and assessing the impact of such mixing on past and future climate change.

First, however, ocean scientists must quantify the subtle ways sound waves veer and bounce as they pass through currents of differing temperature and salinity. Researchers plan to discuss results of their calibration cruises at this week's European Geosciences Union (EGU) meeting in Vienna. W. Steven Holbrook of the University of Wyoming in Laramie, whose team introduced seismic profiling as an oceanography tool in 2003, says he hopes EGU will provide "a real 'meeting of the minds' between seismologists and oceanographers."

To map ocean structure, oceanographers traditionally slow their ship to a crawl, lower instruments every 10 kilometers or so, and interpolate the data points. They have also tracked the spread of dyes, measured surface temperature by satellite, and anchored buoys for long-term observations of ocean currents. In 2000, Holbrook and his team began profiling the ocean with sound. By timing faint echoes from an array of seismic air guns towed behind a ship, they created sub-10-meter-resolution pictures of different water layers across large swaths of sea (*Science*, 8 August 2003, p. 821).

Oceanographers admired the pictures but challenged the geophysicists to put numbers on them. Since 2003, a handful of research cruises, including Holbrook's current U.S. National Science Foundation–funded efforts off Costa Rica, have sought to do that by combining seismic profiling with traditional oceanography. Sheen's 2007 voyage, for example, was part of the European Union–funded Geophysical Oceanography (GO) project led by Richard Hobbs of Durham University in the U.K. GO researchers tested new instruments and techniques aimed at collecting "a definite calibration data set," says Hobbs. One ship

dropped instruments as often as every 2 kilometers behind an air gun–towing ship to obtain more detailed oceanographic data than have historically accompanied seismic profiles. Riffing on the geologists' "ground truth," John Huthnance of the Proudman Oceanographic Laboratory in Liverpool, U.K., says the GO project has given "sea truth to the seismic data."

Once seismic profiling has been fully calibrated, researchers say, terabytes of seismic records from past oil exploration will become ripe for reanalysis. Oil companies stored seismic data from the ocean, which was so much noise to them, in order to subtract it from the much stronger reflection profiles they made of the solid sea floor. "We can happily plug away at legacy data," says White. Mining old data would also bypass the enormous costs of new voyages, which Stephen Jones of Trinity College Dublin in Ireland says cost upward of $25,000 a day.

Legacy data sets aren't perfect. Until the late 1980s, hydrocarbon exploration was largely limited to shallow continental shelves, whereas oceanographers and climate researchers are most interested in ocean-mixing hot spots in deeper waters and at bottlenecks such as the Drake Passage and the Strait of Gibraltar. The biggest limitation on the legacy data is that the petroleum geologists who collected them didn't take enough oceanographic measurements. "There are lots of seismic reflection profiling sections available, but few of them have even a single temperature profile to tell us about the water column," says Raymond Schmitt of Woods Hole Oceanographic Institution in Massachusetts.

Even without completing all their calibrations, oceanographers have published quantitative studies of ocean mixing and in Vienna will discuss engaging eddies with sound. Aided by such progress, they are slowly persuading funding bodies to support seismic profiling despite initially "having some difficulty," says White; the U.K.'s Natural Environment Research Council rejected his and Hobbs's proposals three times. "It's natural to be hesitant to spend money on something which is a little bit unknown," says U.S. Navy oceanographer Warren Wood, a collaborator of Holbrook's, who has obtained internal funding from the Navy.

But Wood says he is "quite impressed by the first results of the GO project." White expects even more later this year, pending analyses by graduate students such as Sheen. For oceanographers, he concludes, seismic profiling "hasn't laid its golden egg, yet."

–LUCAS LAURSEN

Lucas Laursen is a freelance writer in Cambridge, U.K.

Sea hear. Seismic reflections from warm, salty water swirl above the stronger reflections from the sea floor.
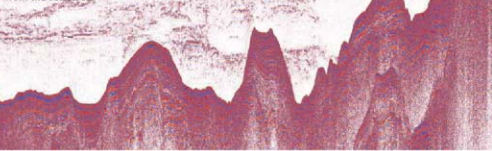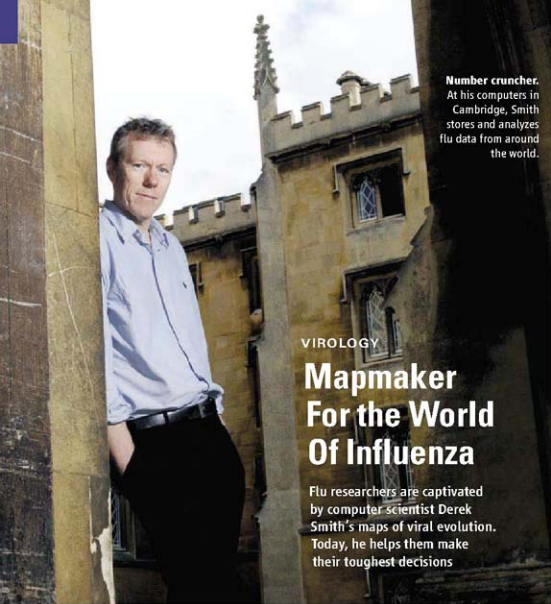
Number cruncher. At his computers in Cambridge, Smith stores and analyzes flu data from around the world.

Mapmaker For the World Of Influenza

Flu researchers are captivated by computer scientist Derek Smith's maps of viral evolution. Today, he helps them make their toughest decisions

DEREK SMITH DIDN'T WANT TO DO ROCKET science—literally. That's how he ended up becoming an internationally recognized expert in influenza virus evolution.

In 1992, at age 33, Smith was working at a research lab of Texas Instruments (TI) in Dallas, a company he had joined a decade earlier, fresh out of a British university. He specialized in the mathematics of speech recognition. One day, a colleague noted that the integrated circuits Smith was developing might play a key role in the control systems for an antiradar missile called HARM that TI was producing for the Pentagon. The missile needed to discern real radar stations from decoys, a problem not unlike detecting subtle differences in spoken words.

"I'm not a pacifist," Smith says, "but I didn't want anything to do with work directly related to the military." Instead, he started looking for a job in which his expertise might benefit public health. He found it in a Ph.D. project to model the immune system's recognition of influenza viruses at the Santa Fe Institute in New Mexico.

He never regretted the choice. Now at the University of Cambridge, U.K., Smith has become the unofficial cartographer of the influenza world. He has developed a technique to produce colorful maps visualizing the never-ending changes in the influenza virus, and over the past 4 years, his lab has become a global nerve center that analyzes influenza data from around the world. His work offers scientists a way to track the virus' evolution almost in real time, says Ian Barr of the World Health Organization's (WHO's) Collaborating Centre for Reference and Research on Influenza in Melbourne, Australia.

Indeed, influenza researchers find Smith's "antigenic cartography" so enlightening that, shortly after he and others published the first results in 2004, he was asked to join the select group that huddles at WHO's headquarters in Geneva, Switzerland, twice a year to decide which strains to put in the annual influenza vaccine that protects 300 million people. "It's a huge responsibility," he says.

From tables to maps

Influenza viruses elude the immune system by changing the shapes of the glycoproteins on their coat—in particular, hemagglutinin (HA), the one that latches onto human cells and to which our immune systems produce antibodies. That's why a flu shot or a natural infection one winter may not protect the year after.

To tell how much a new strain differs from previous ones, researchers test how well its HA is inhibited by antibodies to known strains harvested from infected ferrets. If the antibodies bind well, the new virus is "antigenically close" to those earlier ones; if they don't, the new strain is more distant. These results are used to create complex tables with thousands of numbers, each describing the outcome of one binding assay; they are impenetrable to all but the most experienced researchers.

Smith wanted to turn the tables into clear, accessible maps. Just as mathematicians can reconstruct a decent map of a country from the distance table in the back of a road atlas, it should be possible to map influenza strains based solely on each strain's antigenic distance from the others, he says. So in 1999, Smith teamed up with Alan Lapedes, a mathematician at Los Alamos National Laboratory in New Mexico, who, with Robert Farber, had laid part of the theoretical groundwork for such maps.

He also struck up a collaboration with Ron Fouchier, a virologist at the Erasmus Medical Center in Rotterdam, the Netherlands. When Fouchier switched from HIV research to influenza in 1998, he, too, had been struck by the opacity of the binding assay tables. "I thought there had to be a better way," says Fouchier. The Rotterdam lab also had 3 decades' worth of data and samples—precisely what was needed to produce a map. Of the three influenza types now circulating in humans, the trio picked H3N2, which changes the fastest and affects the most people.

The project was a gamble, says Smith; several groups tried before but failed to get the mathematics right. Even after they produced the first maps, the researchers spent several more years checking the results before publishing.

In the 16 July 2004 issue of *Science* (p. 371), they finally published a map of 273 virus strains that had been isolated since H3N2 emerged in 1968. The map is like that of an archipelago, and the strains come in clusters: Very often, the virus changes little from one year to the next, but occasionally, it makes a major antigenic jump, starting a new cluster, for which existing vaccines offer no protection. The jumps can't always be predicted from the viruses' genetic sequence, because a small change in the *HA* gene can sometimes cause a major shape change in the glycoprotein that makes antibodies lose their grip.

Smith had already presented his results to

Coming Out of Asia—Year In, Year Out

Where does the flu virus hide when there's no flu? That question has puzzled epidemiologists for decades. Every place on Earth has an influenza season, usually the winter, when conditions are best for its spread. But what happens after that? Does the virus lurk in a few people until next year? Or does it disappear and come back, and if so, where from?

Using data about some 13,000 seasonal flu samples from around the world, Derek Smith of Cambridge University in the U.K. and colleagues provide an answer in this issue of *Science* (p. 340): A small number of countries in East and Southeast Asia "seed" the yearly epidemics washing over the planet. "It's really a fantastic paper," says Keiji Fukuda of the World Health Organization (WHO) in Geneva, Switzerland. It shows that strengthening surveillance in Asia is crucial, Fukuda says.

There are plenty of theories on what happens during influenza's absence. Some believed the virus remained in every country, hiding in infected but symptom-free people, or is passed on at rates too low to detect, only to roar back when winter comes around. Others believed it vanished, moving back between the northern and southern hemispheres, for instance, or receding temporarily into tropical Asia, Africa, and South America.

For the new study, Smith and his colleague Colin Russell first analyzed an antigenic map (see main text) of some 13,000 samples of H3N2, the most important flu type currently circulating. They discovered that

changes in the virus always occur first in countries in East and Southeast Asia. That doesn't necessarily mean that the area acts as a source; the virus might also be evolving in parallel around the globe, with Asia being ahead of the curve by a couple of months.

But an analysis of the strains' hemagglutinin genes showed that flu epidemics in Europe, North America, and Australia are actually seeded by countries such as Japan, Thailand, South Korea, and Singapore. Europe and North America then act as conduits to South America, which has less direct contact with Asia.

A study by Edward Holmes of Pennsylvania State University in State College and colleagues, published online by *Nature* this week, also shows that yearly waves in the temperate regions originate in the tropics. But that paper—based on a whole-genome analysis of 1302 strains from New York and New Zealand—does not pinpoint the source.

So what makes East and Southeast Asia special? A variety of climate zones in a small area creates a network of countries with overlapping flu seasons, Smith says. Frequent human travel gives the virus a chance to jump from one country to another. When winter arrives in Europe and the United States, strains from the Asian network spread to those continents aboard jumbo jets. But further, fine-grained studies will be needed to clarify exactly how the Asian network works and whether a similar network exists in India, as Smith and Russell hope to find out together with Indian scientists.

–M.E.

---

some members of the WHO panel; the invitation to join the group followed just 2 weeks after the *Science* paper was published. "All of us could see this was an emerging technology which had immediate application for the work we were doing," says Barr, a member of the group.

Smith's maps increase the group members' confidence that they're making the right choice, says WHO influenza expert Keiji Fukuda; they're also helpful for those less familiar with the tables, he says, such as vaccine producers, regulatory officials, and scientists who don't specialize in influenza. "They can look at these maps and go: 'Oh, now I understand it,'" Fukuda says. Barr concedes that the math and computer wizardry Smith uses to produce his colorful maps go over most influenza scientists' heads. "It will take a while before we throw away our tables," he says.

Sister labs

Meanwhile, Smith's career has taken off. The Univer-

sity of Cambridge made him a research associate in 2003 and a full professor in 2007. In 2005, he landed a $2.5 million U.S. National Institutes of Health Director Pioneers Award

Influenza archipelago. On an antigenic map, each virus strain isolated between 1968 and 2003 appears as a small blob. They occur in clusters (each given a different color), starting with Hong Kong '68. The scale bar represents one antigenic unit, a measure of how similar strains are.

that enabled him to expand his research group to 10 members. He still spends a day a week at Fouchier's lab in Rotterdam, which he says helps keep him grounded in real biology. Some of their best ideas bubble up while the two indulge their shared love of Tom Waits and good whiskey, says Fouchier. Grad students and postdocs, too, are encouraged to cross the North Sea frequently. "We're really like sister labs," Smith says.

"Derek has a wonderful personality for bringing together people and data," says Nancy Cox of the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, one of the four WHO Collaborating Centers. Smith's close association with the Collaborating Centers has given him access to an unparalleled wealth of antigenic and genetic data from around the world,

which enables him to study questions on a global scale, as a paper in this week's issue of *Science* (see sidebar, above) illustrates.

Aware of his privileged position, Smith is careful not to hog the glory, stressing the collaborative nature of the process and crediting the people who provide the data. "As a theoretical biologist, you have to be aware of your place in the food chain," he says. "I don't even know how to do a binding assay."

Now that his lab has come into its own, Smith hopes to tackle new problems. He would like to predict farther in advance which strains will be dominant in a given year. Currently, vaccine producers have just 8 months between the panel's decision and the start of the vaccination season, which means a yearly scramble. At the same time, Fouchier and Smith are trying to predict a strain's antigenic profile directly from its gene sequence; that might eliminate the need for those pesky tables altogether.

Smith also wants to expand the scope of his cartography. Maps for H1N1—the other influenza A virus circulating among humans—and for influenza B are under way. He has also started collaborations to work on agents such as rabies, malaria, and dengue and has plans to branch out into HIV. "There's no reason you can't do the same thing with many other pathogens," he says. One thing seems sure: The mapmaker has put himself firmly on the map.

–MARTIN ENSERINK
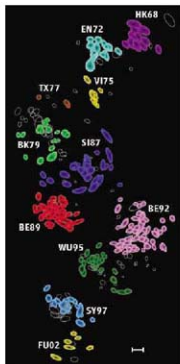
Crystal clear. In an optical lattice, spots of laser light simulate the ions in a crystal. Atoms (red) hopping between the spots simulate the electron in the solid.
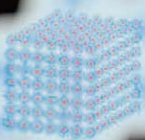
CONDENSED-MATTER PHYSICS

The Mad Dash to Make Light Crystals

Simulations fashioned from laser light and wisps of ultracold atoms might crack the hardest problems in the physics of solids. DARPA wants them in just over a year

For more than a century, physicists have developed ever more sophisticated theories of rock-hard solids and the electrons whizzing within them. They've deciphered metals and insulators, concocted the semiconductors that make computers hum, and explained mind-boggling phenomena such as conventional superconductivity, in which some alloys conduct electricity with no resistance at temperatures near absolute zero.

Yet many problems continue to stump theorists. For example, 22 years after high-temperature superconductors were discovered, physicists still don't know how the exotic compounds carry current without resistance at temperatures up to 138 kelvin. Generally, whenever the shoving among electrons grows too strong, physicists find themselves stymied—even if they resort to high-powered numerical simulations. "We know that we can't do these things on a classical computer," says David Ceperley, a theorist at the University of Illinois, Urbana-Champaign (UIUC).

Help may be on the way, and from an unlikely quarter. Atomic physicists have spent 2 decades fiddling with ultracold gases a millionth the density of air. Now they're striving to model weighty crystalline solids with laser light and cold atoms. Interfering laser beams create an array of bright spots called an "optical lattice" that emulates the ions in a crystal; atoms hopping between the spots emulate the electrons. Physicists can tune the lattice's geometry, the rate of hopping, and the push and pull between atoms. So they hope to map the various behaviors of a model solid—super-

conducting, insulating, and so on—in a portrait called a "phase diagram."

The U.S. Defense Advanced Research Projects Agency (DARPA) has launched a multimillion-dollar program to develop such optical-lattice emulators. They might crack some of the toughest problems in condensed-matter physics or even enable researchers to design materials from scratch. Plans call for the first ones to be running in 15 months.

"The DARPA program is excellent," says Wolfgang Ketterle, an experimenter at the Massachusetts Institute of Technology (MIT) in Cambridge and leader of one of three multi-institution teams receiving funding. "It puts money and resources into an effort that is scientifically superb." But some warn that making the emulators work may be harder than expected. And researchers' goals don't necessarily jibe with DARPA's. Physicists want the phase diagrams. Seeking a tool to design exotic new materials, DARPA wants an automated system that works in just 10 hours.

Abstractions made material

The emulators could bridge the gap between the abstraction of theory and idiosyncrasies of experiments with real solids. Theorists make an educated guess at the physics behind a material's behavior. This "model" is captured in a mathematical expression known as a Hamiltonian, which describes the system's energy. Unfortunately, it's often impossible to "solve" the Hamiltonian to prove it produces the observed behavior. And there is no guarantee that the model doesn't leave out some key detail.

Take, for example, a high-temperature superconductor. It contains planes of copper and oxygen ions arranged in a square pattern along which the electrons pair and glide. At low enough temperatures, the electrons repel one another so strongly they get stuck one-to-a-copper-ion in a traffic jam known as a Mott insulator state. The electrons also act like little magnets, and neighboring electrons point alternately up and down to form an "antiferromagnet." Now, take out a few electrons by tweaking the material's composition. The traffic jam breaks and, perhaps through waves of magnetism, the electrons pair and flow without resistance. Or so many theorists assume.

This scenario is known as the two-dimensional (2D) Fermi-Hubbard model, and nobody can prove it produces superconductivity. Nobody is sure that it captures the essential physics of the messy crystals, either. "The materials are so complicated that you can't look at just the electron-electron correlations," UIUC's Ceperley says. "There are all these other things going on."

But physicists might be able to make a Fermi-Hubbard model by loading cold atoms into a 2D optical lattice that would simulate just these copper-and-oxygen planes. Atoms spinning in opposite directions would hop from bright spot to bright spot. By tweaking the laser beams and applying a magnetic field, physicists would vary the rate of hopping, the repulsion between atoms, and other factors to determine under what conditions if any the model produces superconductivity, says Tin-Lun "Jason" Ho, a theorist at Ohio State University in Columbus. "The goal is to reproduce the model faithfully in an optical lattice and let nature tell you what the solution is," he says.

The next coolest thing

The push marks the next chapter in the short, glorious history of ultracold atoms. Atoms can be sorted into two types—bosons and fermions—depending on how much they spin. Thanks to quantum mechanics, the two types behave very differently. Bosons are inherently gregarious. In 1995, two teams independently chilled bosons below a mil-

lionth of a kelvin to coax them into a single quantum wave and produce a state of matter called a Bose-Einstein condensate (BEC) that flows without resistance (*Science*, 14 July 1995, p. 152). That accomplishment netted a Nobel Prize in 2001.

Fermions are loners, as no two identical fermions can occupy the same quantum wave or state. Nevertheless, at very low temperatures fermions can get it together to flow freely. First they have to pair, and then the pairs condense into a quantum wave. This is what happens in superconductors, and in 2004, physicists made fermionic atoms pair and condense in much the same way (*Science*, 6 February 2004, p. 741).

Given those accomplishments, creating an optical-lattice emulator might seem easy. Electrons are also fermions, so it might appear that researchers need only impose an optical lattice on fermionic atoms already trapped by magnetic fields and laser beams. But researchers have several steps to go before they can emulate the Fermi-Hubbard model and other intractable systems. They must achieve the Mott insulator state, which would pin one atom to each lattice site, and then the antiferromagnetic state, in which neighboring atoms spin in different ways.

Physicists have made progress. In 2002, Immanuel Bloch, now at the Johannes Gutenberg University of Mainz in Germany, and colleagues reached the Mott insulator state for bosons by loading a BEC of rubidium-87 into an optical lattice and cranking up the brightness of the laser spots to effectively increase the repulsion between atoms. "That was a landmark," says Randall Hulet, an experimenter at Rice University in Houston, Texas. "That showed that we could do something that was relevant from a condensed-matter perspective."

Last month at an American Physical Society meeting in New Orleans, Louisiana, Niels Strohmaier and Tilman Esslinger of the Swiss Federal Institute of Technology Zurich reported reaching the more elusive Mott state for fermions. "We have a few puzzle pieces, and now we want to put everything together," says Ketterle, who shared the Nobel for BECs.

Emulators, pronto!

The DARPA program aims to do just that. Last July, the agency gave three large teams—led by Ketterle, Hulet, and Christopher Monroe at the University of Maryland, College Park—a few million dollars each (DARPA won't say exactly how much) and 2 years to develop a working emulator. In that first phase, researchers will tackle simpler models for

which the Hamiltonian *can* be solved. For example, Hulet's group will study fermions in 1D tubes of light, and Ketterle will aim for the antiferromagnetic state of fermions in a 3D lattice.

If a team's starter emulation works by July 2009, it will be eligible for a 3-year second phase, in which researchers will tackle an incalculable Hamiltonian. Hulet and Ketterle both hope to emulate superconductivity in the 2D Fermi-Hubbard model. Monroe's team is focusing on bosons in both phases, which do not mimic electrons but should still be useful for simulating exotic magnetic materials.

To get to the second phase, the emulators for the first phase must work at lightning speed, however. The machinery must step through a complete phase diagram in 10 hours, not including the setup time. That's roughly how long it takes the best computer simulations to run, says DARPA program manager Air Force Lt. Col. John Lowell. "You're trying to establish a comparison with other computational techniques, and time is the metric," he says.

It sounds like the sort of results-on-demand program that would drive university researchers crazy. However, all voice great enthusiasm for the project. "DARPA really wants you to stay focused on the task at hand, and I find that very productive," Hulet says. "I've got a schedule on my white board of when things have to get done, and it definitely creates some tension in the lab."

The challenges ahead

Making the emulators work won't be easy, physicists say. The biggest hurdle may be getting the atoms cold enough. Researchers may have to chill gases to picokelvin temperatures to emulate the Fermi-Hubbard model. Oddly, they may catch a break getting part of the way there. Theory suggests that if they turn on an optical lattice gently—so as not to add entropy—a gas of fermions should spontaneously cool enough to reach the antiferromagnetic state. But researchers may be underestimating the difficulty of getting even colder to reach the superconducting state, Ohio State's Ho says. "It requires a breakthrough," he says. "Just doing things the way they are doing them now is as good as praying."

Experimenters will also have to devise ways to prove that their emulator is doing what they think it is. A high-temperature superconductor may be messy, but it produces an unambiguous signal that it's working: zero electrical resistance. Atoms in a lattice won't signal so clearly that they have gone superconducting, so proving they have will require subtle new probes.

Then there is the 10-hour time limit—an odd requirement given that physicists would be happy to have the phase diagram for the Fermi-Hubbard model even if it took years to get it. Some predict DARPA officials won't enforce the limit strictly. "Everybody knows it can be tweaked in the end if need be," Monroe says. Don't be so sure, Lowell warns. "I wouldn't have laid this out as a milestone if I didn't think it was doable," he says. "And you wouldn't have signed on to it if you didn't think it was achievable."

Where will it all lead? Even leading physicists doubt that they'll produce a black box capable of deciphering any solid. "I see it as unlikely that in the end there will be this universal machine that solves any problem you would like to solve," says Mainz's Bloch. Some say optical lattices may serve primarily to validate techniques for computer simulation, which will remain the biggest wrench in the theorist's toolbox.

Nevertheless, all agree that making light crystals could be a revolutionary advance. "It's bloody difficult, but it doesn't seem impossible," Ketterle says. "Let's stop talking and start doing." The clock is already running.

–ADRIAN CHO

Heavy hitters. MIT's Wolfgang Ketterle (*top*) and Rice's Randall Hulet praise DARPA's vision.

LETTERS

edited by Jennifer Sills

Coral Adaptation in the Face of Climate Change

IN THEIR REVIEW, "CORAL REEFS UNDER RAPID CLIMATE CHANGE and ocean acidification" (14 December 2007, p. 1737), O. Hoegh-Guldberg *et al.* present future reef scenarios that range from coral-dominated communities to rapidly eroding rubble banks. Notably, none of their scenarios considers the capacity for corals to adapt. The authors dismiss adaptation because "[r]eef-building corals have relatively long generation times and low genetic diversity, making for slow rates of adaptation [relative to rates of change]." We think the possibility of adaptation deserves a second look.

Many features of coral life histories, such as extended life spans, delayed maturation, and colony fission, do result in long generation times (*1*) [some between 33 and 37 years (*2*)]. However, other corals, such as many species of *Acropora* and *Pocillopora*, mature early, grow rapidly, and suffer whole-colony mortality, as opposed to colony fission, after mechanical disturbances (*3*) and thermal stress (*4*). The life histories of these ecologically important and abundant species suggest an underappreciated capacity to adapt rapidly to changing environments.

Repeated bleaching episodes in the same coral assemblages and the increasing scale and frequency of coral bleaching have been cited as evidence that corals have exhausted their genetic capacity to adapt to rising sea surface temperatures (*5*). However, comparisons of the rates of mortality within populations among bleaching events are not available. Without these data, it is not possible to assess whether the adaptive response has been exhausted. Indeed, the effects of temperature and acidification on even the most basic vital rates in corals, such as

growth, mortality, and fecundity, are largely unknown, as are the physiological trade-offs among these traits. Consequently, the sensitivity of population growth to climate-induced changes in vital rates remains almost completely unexplored [but see (*6*)]. In the absence of long-term demographic studies to detect temporal trends in life history traits, predicting rates of adaptation, and whether they will be exceeded by rates of environmental change, is pure speculation. Indeed, where such data are available for terrestrial organisms they demonstrate that contemporary evolution in response to climate change is possible (*7*).

ANDREW BAIRD[1] AND JEFFREY A. MAYNARD[2]

[1]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia. [2]Centre of Excellence for Environmental Risk Analysis, University of Melbourne, Parkville, VIC 3010, Australia.

References
1. T. P. Hughes, D. Ayre, J. H. Connell, *Trends Ecol. Evol.* **7**, 292 (1992).
2. R. C. Babcock, *Ecol. Monogr.* **61**, 225 (1991).
3. J. S. Madin, S. R. Connolly, *Nature* **444**, 477 (2006).
4. A. H. Baird, P. A. Marshall, *Mar. Ecol. Prog. Ser.* **237**, 133 (2002).
5. O. Hoegh-Guldberg, *Mar. Freshwater Res.* **50**, 839 (1999).
6. M. Wakeford, T. J. Done, C. R. Johnson, *Coral Reefs* **27**, 1 (2008).
7. D. K. Skelly *et al.*, *Conserv. Biol.* **21**, 1353 (2007).

Response

WE CERTAINLY HOPE THAT BAIRD AND MAYnard are right and that in the coming years corals will exhibit an adaptive capability that they have not yet exhibited in situ or in the laboratory. At this point, however, it appears unlikely.

As Baird and Maynard point out, the coral genera *Acropora* and *Pocillopora* have generation times that are short (several years) relative to the generation times of other corals. The majority of coral generation times, however, are still long (decades) relative to the accelerating pace of climate change, throwing doubt on the scope of most coral species for

rapid adaptation (*1*).

Corals, like other organisms, can also modify the risk of coral bleaching over the short term through physiological acclimation (*2*). Acclimation, however, as with any phenotypic change, is limited. In the same vein, corals that form symbioses with more than one variety of dinoflagellate can shift their populations so that they are dominated by their more thermally tolerant dinoflagellate genotypes during thermal stress. Unfortunately, these short-lived changes have not yet resulted in the novel host-symbiont combina-

tions that will be required for survival in the challenging temperatures and acidities of future oceans under rising atmospheric carbon dioxide.

It is important not to confuse genetic adaptation with the increased average thermal tolerance observed for some coral communities over the past 25 years, which has occurred largely because thermally sensitive species have died out, leaving robust species behind

(3). Equally important is the lack of evidence that corals have the capacity to either acclimate or adapt to falling aragonite saturation states. It seems unlikely that genetic adaptation will solve the problems of global change facing corals. Indeed, paleontological evidence indicates that calcifying marine organisms including corals suffered a protracted period of absence after large and rapid changes in atmospheric carbon dioxide associated with the Permian–Triassic extinction event *(4, 5)*. It took millions of years for these organisms and ecosystems to recover.

O. HOEGH-GULDBERG,[1]* P. J. MUMBY,[2] A. J. HOOTEN,[3] R. S. STENECK,[4] P. GREENFIELD,[5] E. GOMEZ,[6] D. R. HARVELL,[7] P. F. SALE,[8] A. J. EDWARDS,[9] K. CALDEIRA,[10] N. KNOWLTON,[11] C. M. EAKIN,[12] R. IGLESIAS-PRIETO,[13] N. MUTHIGA,[14] R. H. BRADBURY,[15] A. DUBI,[16] M. E. HATZIOLOS[17]

[1]Centre for Marine Studies, The University of Queensland, St Lucia, QLD 4072, Australia. [2]Marine Spatial Ecology Laboratory, School of Biological and Chemical Sciences, University of Exeter, Exeter EX4 4PS, UK. [3]AJH Environmental Services, 4900 Auburn Avenue, Suite 201, Bethesda, MD 20814, USA. [4]School of Marine Sciences, Darling Marine Center, University of Maine, Walpole, ME 04573, USA. [5]The Chancellery, University of Queensland, St Lucia, QLD 4072, Australia. [6]Marine Science Institute, University of the Philippines, Diliman, Quezon City, Philippines. [7]Department of Ecology and Evolutionary Biology, E321 Corson Hall, Cornell University, Ithaca, NY 14853, USA. [8]International Network on Water, Environment and Health, United Nations University, Hamilton, ON L8N 1E9, Canada. [9]School of Biology, Ridley Building, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK. [10]Department of Global Ecology, Carnegie Institution of Washington, Stanford, CA 94305, USA. [11]National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA. [12]NOAA Coral Reef Watch, NOAA E/RA31, Silver Spring, MD 20910–3226, USA. [13]Unidad Académica Puerto Morelos, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México, Cancún, 77500 QR, México. [14]Wildlife Conservation Society, Bronx, NY 10460, USA. [15]Resource Management in Asia-Pacific Program, Australian National University, Canberra ACT 0200, Australia. [16]Institute of Marine Sciences, University of Dar es Salaam, Tanzania. [17]Environment Department, MC5-512, The World Bank, Washington, DC 20433, USA.

*To whom correspondence should be addressed. E-mail: oveh@uq.edu.au

References

1. D. K. Skelly *et al.*, *Conserv. Biol.* **21**, 1353 (2006).
2. S. L. Coles, B. E. Brown, *Mar. Biol.* **46**, 152 (2003).
3. P. W. Glynn, J. L. Maté, A. C. Baker, M. O. Calderón, *Bull. Mar. Sci.* **69**, 79 (2001).
4. G. D. Stanley Jr., *Earth-Sci. Rev.* **60**, 195 (2003).
5. R. S. Steneck, *Paleobiology* **9**, 44 (1983).

Freshwater Forcing: Will History Repeat Itself?

IN THEIR RESEARCH ARTICLE "REDUCED North Atlantic deep water coeval with the glacial Lake Agassiz freshwater outburst" (4 January, p. 60), H. F. Kleiven *et al.* present compelling evidence for an abrupt deep-ocean response to the release of fresh-

water from glacial Lake Agassiz into the northwest Atlantic about 8400 years ago. Such data are particularly important in evaluating the response in ocean models of the Atlantic Meridional Overturning Circulation (MOC) to freshwater forcing. For this event, the freshwater forcing was likely large but short; Clarke *et al.* *(1)* estimate that the flood had a freshwater flux of 4 to 9 Sv released in 0.5 years.

In this context, we are aware of no possible mechanism that might reproduce such a forcing in response to global warming, and all available model simulations, including those with estimates of maximum Greenland Ice Sheet (GIS) melting rates, indicate that it is very unlikely that the MOC will undergo an abrupt transition during the course of the 21st century *(2)*. Multimodel ensemble averages under Special Report on Emissions Scenario (SRES) A1B suggest a best estimate of 25 to 30% reduction in the overall MOC strength *(2)*. In one example, 14 coupled models simulated a 100-year 0.1-Sv freshwater perturbation to the northern North Atlantic Ocean—17 times the recently estimated melt rates from the GIS—and the MOC weakened by a multimodel mean of 30% after 100 years; none of the models simulated a shutdown *(3)*. Another model simulated greenhouse gas levels that increased to four times preindustrial values and then remained fixed; the resulting GIS displayed a peak melting rate of about 0.1 Sv, with little effect on the MOC *(4)*. One model simulation uses the SRES A1B scenario but adds an additional 0.09-Sv

freshwater forcing as an upper-bound estimate of potential GIS melting. In this case, the MOC weakened but subsequently recovered its strength, indicating that GIS melting would not cause abrupt climate change in the 21st century *(5)*. Accordingly, we urge caution in drawing comparisons of the abrupt change 8400 years ago to future scenarios involving, for example, the melting of the GIS and its relevance to human societies.

PETER U. CLARK,[1] THOMAS L. DELWORTH,[2] ANDREW J. WEAVER[3]

[1]Department of Geosciences, Oregon State University, Corvallis, OR 97331, USA. [2]Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, NJ 08542, USA. [3]School of Earth and Ocean Sciences, University of Victoria, Victoria, BC V8W 3P6, Canada.

References

1. G. K. C. Clarke, D. W. Leverington, J. T. Teller, A. S. Dyke, *Quat. Sci. Rev.* **23**, 389 (2004).
2. G. A Meehl *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, New York, 2007), pp. 747–845.
3. R. J. Stouffer *et al.*, *J. Clim.* **19**, 1365 (2006).
4. J. K. Ridley, P. Huybrechts, J. M. Gregory, J. A. Lowe, *J. Clim.* **17**, 3409 (2005).
5. J. H. Jungclaus, H. Haak, M. Esch, E. Roeckner, J. Marotzke, *Geophys. Res. Lett.* **33**, 10.1029/2006GL026815 (2006).

Response

WE THANK CLARK *ET AL.* FOR REITERATING AN important point regarding the relevance of our study (4 January, p. 60) for future global warming scenarios. We agree with Clark and colleagues that the 8400-year deep circulation anomaly we reported, although useful for evaluating the response of ocean models

to sudden fluxes of freshwater, does not represent the most realistic (one-to-one) analog for possible future changes. Indeed, we found that only one such extreme deep circulation anomaly occurred in the Holocene and that it followed the rapid drainage of an enormous proglacial lake, for which we also know of no foreseeable equivalent in our future. In addition, we pointed out that the ocean circulation prior to the outburst flood was most likely different than it is today—Labrador sea convection and Danish Straight Overflow Water were both thought to be weaker than today (1, 2). Finally, our records

demonstrate just how complex the relationship between climate and ocean circulation was during the rest of the Holocene.

We demonstrated that the ocean sensitively responded to the extreme freshwater forcing event ~8400 years ago. Our results agree with modeling studies applying similarly large freshwater fluxes, confirming that the deep ocean can change just as quickly as models predict (3). In the most general sense, this supports the idea that the estimated 25 to 30% reduction (4) in Meridional Overturning Circulation (MOC) referred to by Clark et al. is plausible on century time scales.

Our approach for understanding the extreme and distinctly different scenario ~8400 years ago may also be useful in determining the sensitivity and thresholds of ocean circulation for the more modest but sustained freshwater forcing expected in our future. Further work will be necessary to validate the scale and rate of MOC changes estimated by models in these intermediary states. A natural next step would be to provide a detailed characterization of deepwater properties and circulation at times in our past that contain elements more in com-

mon with our future. One obvious candidate is the previous interglacial period (Marine Isotope Stage 5e), which was warmer than the present (5), had a smaller Greenland Ice Sheet, and may have experienced a sea-level rise at a similar rate to that projected (6).

H. F. KLEIVEN,[1] CATHERINE KISSEL,[2] CARLO LAJ,[2] ULYSSES S. NINNEMANN,[3] THOMAS O. RICHTER,[4] ELSA CORTIJO[2]

[1]Bjerknes Centre for Climate Research and Department of Earth Science, University of Bergen, Allegaten 41, 5007 Bergen, Norway. [2]Laboratoire des Sciences du Climat et de l'Environnement, Avenue de la Terrasse, Domaine du CNRS, Gif-sur-Yvette cedex, 91198, France. [3]Department of Earth Science and Bjerknes Centre for Climate Research, Allegaten 41, 5007 Bergen, Norway. [4]Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, Netherlands.

References
1. C. Hillaire-Marcel, A. de Vernal, D. J. W. Piper, Geophys. Res. Lett. 34, L15601 (2007).
2. C. Hillaire-Marcel, A. de Vernal, G. Bilodeau, A. J. Weaver, Nature 410, 1073 (2001).
3. A. P. Wiersma, H. Renssen, H. Goose, T. Fichefet, Clim. Dyn. 27, 831 (2006).
4. G. A. Meehl et al., in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, S. Solomon et al., Eds. (Cambridge Univ. Press, New York, 2007), pp. 747–845.
5. B. L. Otto-Bliesner et al., Science 311, 1751 (2006).
6. E. J. Rohling et al., Nat. Geosci. 1, 38 (2008).

ECONOMICS

First, Kill the Economists

E. Roy Weintraub

The prophet Jeremiah is alive and well and teaching economics at Harvard. It is not often that a scholar with no particular historical or philosophical expertise trashes the Western enlightenment in order to stomp on the discipline of economics as a manifestation of all that was lost in creating the modern world.

Stephen A. Marglin's argument in *The Dismal Science* is that economics—with its focus on an individual's preferences, the freedom to engage in activities to promote his or her well-being, and the pursuit of self-interest variously construed—perverts a natural moral order: "the foundational assumptions of economics are in my view simply the tacit assumptions of modernity. The centerpiece in both is the rational, calculating, self-interested individual with unlimited wants for whom society is the nation-state." And what modernity shunned was "community."

His main line is that "The market undermines community because it replaces personal ties of economic necessity by impersonal market transactions.... The ambivalent relationship between noneconomists and economics reflects the ambivalence with which modernity is regarded." To be sure, sociologists deal with community, as do anthropologists, as do political scientists, and so on. But economics, for Marglin, is different: "Economics is not only descriptive; it is not only evaluative; it is at the same time constructive—economists seek to fashion a world in the image of economic theory." Economics and thinking like an economist are bad for the health of the world. Indeed, he closes his volume stating that "There are many ways of resolving the tensions between individualism and holism, between self-interest and obligation to others, between algorithm and experience, between the claims of various communities on our allegiance, between material prosperity and spiritual health. Economics offers one way, but as presently constituted, economics is hobbled

by an ideology in which these tensions are replaced by a set of pseudo-universals about human nature. A dismal science indeed."

The argument about the proper way to do economics is an old one. An 1832 complaint in *The Eclectic Review* charged the work of Thomas Malthus and David Ricardo with leading the public far from "the true path of inquiry" and making political economy "a hideous chain of paradoxes at apparent war with religion and humanity." In the past century or two, we have heard this lamentation from time to time from both secular and religious figures.

In much of Europe, what we now call economics developed in order to understand various matters of business law, contracts, taxation, international trade, and project management. Issues like tariff policy and currency management were discussed by individuals who were variously lawyers, engineers, politicians, managers, and business people, and training in such expertise developed pari passu.

The professionalization of economics was a late 19th century phenomenon. Cambridge's Alfred Marshall, in attempting to construct a scientific economics, was not able to establish economics as a separate discipline until the death of Henry Sidgwick, the university's professor of moral philosophy, under whose direction lectures in political economy had been organized. In the United States at that time, economics was growing from different sources. One stream followed from individuals who had obtained Ph.D.'s in Germany, where social policy issues—labor unions, socialism, the nascent welfare state, etc.—were galvanizing the universities. But a second stream nurturing the American progressive economists grew from the social gospel movement, which sought to promote the kingdom of God on Earth through enlightened social policy and the kind of market interven-

tions that Adam Smith in fact quite welcomed.

The kind of economics from which Marglin recoils is, however, not of the sort that was present in writings of individuals (e.g., Smith, Ricardo, John Stuart Mill, Marshall, and John Commons) who have been claimed as ancestors by modern economists. It is instead what developed in the post–World War II stabilization of economic discourse and the final professionalization of the discipline. It was during that postwar period, not in the Enlightenment, that economic science became normal in Thomas Kuhn's sense.

Marglin's account appears confused by this history. Moreover, he appears to believe that the ideas he engages and then casts aside (ideas about the economic agent, preferences, equilibrium, models, and markets) all grew up not in the 20th century but hundreds of years earlier—and that those ideas have had stable meanings ever since: "For four hundred years, economists have been active in the enterprise of constructing the modern economy and society, both by legitimizing the market and by promoting the values, attitudes, and behaviors that make for economic success. No apology is due for this—except for the pretense of scientific detachment and neutrality and the unwillingness to confront the ideological beam in our collective eye." The ahistoricity of such a statement is startling; for instance, it assumes wrongly that there were individuals called economists 400 years ago and that science in 1600 meant the same thing as it does in 2008.

In his critique, Marglin moves back and forth between moralizing about the loss of community and contempt for the economists' tools and models. He claims, "By promoting market relationships, economics undermines reciprocity, altruism, and mutual obligation, and therewith the necessity of community. The very foundations of economics, by justifying the expansion of markets, lead inexorably to the weakening of community." He complains that "it is difficult to tell a plausible story of how individuals acquire meaningful preferences between consumption today and consumption a decade or two hence, in the way one can imagine learning about peaches today and pears today." But is not Marglin's Harvard College teaching an instruction of the young designed to shape their preferences, especially preferences about long-term versus short-term goals?

From the first times economic arguments were parsed and markets described, there were those who found both contemptible, and this

The Reverend Thomas R. Malthus.

The reviewer is at the Department of Economics, Duke University, Durham, NC 27708–0097, USA. E-mail: erw@econ.duke.edu

was well before the Enlightenment. Attacks on money lending at interest go back even earlier than Jesus on the temple steps. Recall Aquinas's ideas about the "just price." One mustn't forget Shakespeare's Shylock, either. Tax collecting for kings and emperors requires economic management skills, but no one likes to pay taxes. In a prize-winning book (*1*), William Coleman showed how over the centuries the very idea of economics has been loathed by left, right, and center; Christian, Jew, and anti-Semite; pope and communist dictator; lawyer and business mogul; and scientist and humanist.

In this same tradition of anti-economics, Marglin sees the future of the field as bleak, with the current generation of economics students avoiding large questions in their search for career advancement. And the problems that economics creates will only get worse, he claims, because globalization will make the national community as obsolete as the market has made the local community.

I note in closing that the lead dust-jacket blurb for this volume was provided by the noted economist and social theorist Bianca Jagger (sic). Whatever was Harvard University Press thinking?

Reference
1. W. O. Coleman, *Economics and Its Enemies: Two Centuries of Anti-Economics* (Palgrave Macmillan, New York, 2002).

10.1126/science.1157088

PHYSIOLOGY

Toward the Dominance of Vision?

Andreas Keller

If we want to know the time, we look at a clock; if we want to find out whether the milk is still fresh, we check the expiration date printed on the carton. Mark M. Smith's *Sensing the Past: Seeing, Hearing, Smelling, Tasting, and Touching in History* reminds us that it has not always been like that. There was a time when the ring of a bell signified the hour and we decided whether the milk had soured by sniffing or tasting it. *Sensing the Past* traces how the relevance of the individual senses has changed through the years. Smith's reading of

The reviewer is at the Laboratory of Neurogenetics and Behavior, Rockefeller University, 1230 York Avenue, Box 63, New York, NY 10065, USA. E-mail: kellera@mail.rockefeller.edu

Fragrant rose. "*Rosa graeca*" (Dutch, late 16th century).

the extensive historical and anthropological literatures leads him to challenge the theory, championed by the pioneers in sensory history (*1*, *2*), that around the time the printing press was invented vision began to replace taste, smell, touch, and hearing. Smith (a historian at the University of South Carolina) argues that nonvisual senses continue to be relevant in the modern world.

Consequently, the book examines the historical and present-day use of each of the five senses in equal depth and stresses the importance of interactions among them. The author collects his examples from different historical periods and includes studies from around the globe, although the dominance of primary literature dealing with Western history makes a truly balanced account impossible.

Some studies support the notion that there is a tendency toward an increasing importance for vision. For example, roses were once primarily prized for their scents. Since the Enlightenment, cultivators have selected roses for their looks, which has led to the big, beautiful, but scentless blooms on long stems that we now buy on Valentine's Day. Other observations support Smith's suspicion that this trend toward vision is not universal. Taste, smell, and touch gained in importance relative to vision in American supermarkets. In place of "don't touch" signs that were ubiquitous in the early 20th century, we now find customers squeezing and sniffing the displayed fruits. The parallel rise of international trade and the importance of taste offers another instance of

a nonvisual sense gaining importance. In Europe, trade introduced sugar, potatoes, coffee, cacao, tea, peppers, and exotic fruits into the diet on a large scale—which led to eating habits no longer based on availability but on, well, taste. With the increase in options, taste was elevated to a sense of social, cultural, and political importance.

Cases like this, in which sensory perception is of social and political significance, make up most of the book. For example, Smith points to the "stench of the working class" and the almost entirely social meaning of the "lowest of the senses" (touch), which is intimately associated with ownership and therefore with power. Slaves could be touched and beaten at will by slaveholders, and men did "push women, rub against, hit, and sexually assault them in private and public." Touching served to reaffirm status and assert authority. The book's most detailed discussions concern slavery and race in the American South, the topics of earlier books by Smith. He retells the fascinating story of a "visually 'white' 'black' man" in Louisiana who was refused a seat in a car for white passengers on the train—the *Plessy* v. *Ferguson* case (1896) through which the U.S. Supreme Court enshrined the "separate but equal" doctrine. Illustrating the arbitrariness of the alleged dominance of vision over the other senses, Louisiana's attorney explained that "I might not be able to see that he is black, but I can certainly smell his racial identity."

With its overview of recent studies, *Sensing the Past* offers an informative and entertaining introduction to the underappreciated field of sensory history. The extensive bibliography makes it a valuable resource for readers who wish to further explore historical changes in the relative importance of the senses. For a neuroscientist like myself, this short book is an important reminder that "sensory perception is a cultural, as well as a physical act." How something looks, smells, feels, tastes, or sounds depends on the physical properties of the stimulus, on the workings of the brain, and on the cultural context.

References
1. M. McLuhan, *The Gutenberg Galaxy: The Making of Typographic Man* (Univ. of Toronto Press, Toronto, 1962).
2. W. J. Ong, *Orality and Literacy: The Technologizing of the Word* (Routledge, New York, 1982).

10.1126/science.1157191

ECOLOGY

Agriculture at a Crossroads

E. Toby Kiers,[1]* Roger R. B. Leakey,[2] Anne-Marie Izac,[3] Jack A. Heinemann,[4] Erika Rosenthal,[5] Dev Nathan,[6] Janice Jiggins[7]

The present path of agricultural development will not achieve development goals according to a recent assessment, but a solid foundation for improvements exists.

Recent scientific assessments (1–4) have alerted the world to the increasing size of agriculture's footprint, including its contribution to climate change and degradation of natural resources (5). By some analyses, agriculture is the single largest threat to biodiversity (6). Agriculture requires more land, water, and human labor than any other industry (7). An estimated 75% of the world's poor and hungry live in rural areas and depend directly or indirectly on agriculture for their livelihoods (8). As grain commodity prices rise and per capita grain production stagnates (9), policy-makers are torn between allocating land to food or fuel needs. The governance of agriculture requires new thinking if it is to meet the needs of humanity now and in the future. The International Assessment of Agricultural Science and Technology for Development (IAASTD) brought together governments, international organizations, and private sector and civil society organizations to address these challenges (10). The task was to assess the current state and future potential of formal and informal knowledge, as well as science and technology (S&T), (i) to reduce hunger and poverty, (ii) to improve rural livelihoods, and (iii) to facilitate equitable, sustainable development.

The IAASTD recently released its assessment (11). The assessment acknowledges the enormous historical contributions of S&T to increased yields, nutrition, and aggregate wealth but also recognizes that gains have been uneven and that successes have been accompanied by environmental and social consequences. Production increases have not consistently improved food access for the world's poor. Where production has been intensified, it has generally been accompanied by costs such as extensive eutrophication from fertilizer runoff, pesticide contamination, and loss of local crop landraces (12). The assessment found that structural changes in governance, development, and delivery of S&T are required so that benefits are shared more equitably and environmental impacts are lessened.

Controversy arising from the assessment's findings (13–15) has focused on a single element of the study, namely, the role of transgenics, particularly genetically modified (GM) crops. The assessment, however, was tasked with appraising the contribution of a diversity of S&T approaches to the combined social, environmental, and production goals. GM technology was not rejected in principle; the assessment found GM crops appropriate in some contexts, unpromising in others, and unproven in many more. The potential of GM crops to serve the needs of the subsistence farmer is recognized, but this potential remains unfulfilled. No conclusive evidence was found that GM crops have so far offered solutions to the broader socioeconomic dilemmas faced by developing countries. Here, we, as IAASTD authors, summarize the wider key actions identified in the assessment and the solutions they offer.

Redirection of agricultural S&T. Inadequate attention has been devoted to the generation, dissemination, and uptake of S&T that addresses the needs of the rural poor and to developing technologies that lessen the environmental impacts of agriculture. A meager one-third (about U.S. $10 billion) of all global research expenditure on agriculture is spent on solving the problems of agriculture in developing countries (16), home to ~80% of the global population. This amount is less than 3% of the total value of agricultural subsidies that countries of the Organization for Economic Cooperation and Development (OECD) pay to maintain their agricultural output (16). Consequently, regions with severe biophysical constraints and marginalized communities have historically benefited least from S&T development (17).

In the next two decades, climate change is predicted to cause major crop losses in the world's poorest regions (18). The driest areas of the world are already home to more than 2 billion people. Agricultural S&T has yet to offer effective rural management options for crop and livestock systems appropriate for water-constrained dry lands and stress conditions. Except for the Consultative Group on International Agricultural Research (CGIAR) (19), few others have sought crop improvements in the small-grain cereals, tubers, and legumes cultivated by hundreds of millions of farmers.

Will private sector companies lead this redirection? There is plenty of scope for them to play a vital role, as they already dominate the research landscape. Private sector investments in agricultural research and development (R&D) reached more than $12 billion in 2000, 30 times the budget of the entire CGIAR international agricultural research system (20). A redirection of S&T is needed to move away from processes that have profited primarily large-scale enterprises to processes that address the most basic needs of the world's 900 million small farmers. The availability and cost of good-quality seed, especially in sub-Saharan Africa, pose real constraints for poor farmers (21), as does severe soil degradation and post-harvest losses.

Approaches	Redirection	Arrangements, laws, regulations
Farmer participation	Generation of S&T	IPR to support farmer innovation
Funding for affordable technology development		
Governance to allow public deliberation of S&T	Policy and planning of S&T	Regional and international forums to drive S&T planning
		Governmental regulation of private sector
Access to trade and market analysis	Access and exchange of S&T	New information and communication tools for rural communities
Funding for higher education		
Access to natural resources	Capacity development	Decentralized R&D facilities
Building local expertise		Rural to urban supply chains
		Research networks

Translating redirection of agricultural S&T into concrete approaches, arrangements, laws, and regulations.

[1]Institute of Ecological Science, Faculty of Earth and Life Sciences, Vrije Universiteit, De Boelelaan 1085, NL-1081 HV Amsterdam, Netherlands. [2]James Cook University, School of Marine and Tropical Biology, Cairns, Queensland, Australia. [3]Alliance of the CGIAR Centres, c/o IFAD, 200 Via del Serafico, Rome, Italy. [4]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. [5]Center for International Environmental Law, 1350 Connecticut Avenue, NW, Suite 1100, Washington, DC, USA. [6]Institute for Human Development, New Delhi, India. [7]Communication and Innovation Studies, Wageningen University Research, Wageningen, Netherlands.

*Author for correspondence. E-mail: ekiers@falw.vu.nl

All of these problems can be tackled with relatively simple technologies and investments. Evolving intellectual property rights (IPR) regimes to encourage farmers' entrepreneurship and initiatives to develop small seed companies can improve delivery of locally appropriate seeds to poor farmers, not currently offered by the few companies dominating the global seed market (22). Reversing soil infertility through use of locally available resources (e.g., nitrogen-fixing trees, indigenous rock phosphate) has increased food security for tens of thousands of African farmers (23). Recent research from the Food and Agriculture Organization of the United Nations (FAO) suggests that total milk spoilage, spillage, etc., in East Africa and the Near East costs small farmers $90 million/year. Dairy imports to the developing world, which increased 43% between 1998 and 2001, could have been significantly reduced with simple on-farm post-harvest technologies (24). Similar investments in affordable technologies (e.g., small metallic silos) could prevent rice post-harvest losses ranging between 8 and 26% in China (25).

There is a need to capitalize on human ingenuity, deployed for centuries to solve agricultural challenges. Scientists at the African Rice Center are adapting the use of golden weaver ants (a centuries-old technology developed by farmers in Asia) as a pest control method, so West African mango producers can access profitable European markets (26). In some cases, existing small-scale farming systems have high water-, nutrient-, and energy-use efficiencies and conserve resources and biodiversity without sacrificing yield. The extrapolation of these principles to larger-scale farming is another critical research direction (2).

Developing S&T to increase agricultural market access for rural communities is needed, including optimizing rural supply chains, increasing local addition of value, and simple, but effective, measures like enhancing market feeder roads. S&T has largely ignored using "wild" species as resource production systems, even though their positive impacts are clear (27). Such initiatives engage communities in decision-making processes while building production capacities.

Innovation. Initiatives in which local communities effectively set the agenda, alongside S&T developers, have emerged in the last decade. Farmers and formal plant breeders in West Africa are creating rice varieties that compete effectively with weeds to relieve labor shortages, alongside dual-purpose cowpea varieties with good yields followed by a green foliage harvest for livestock (28), and farmer-led seed multiplication strategies for stressful climate and economic conditions

(19). Examples from fisheries, rural energy, and agro-processing all abound. It is not the technologies that are innovative here, but the pathway to their development, which involves continuous on-site cycles of learning and change (27).

The assessment's message is clear: Innovation is more than invention. Success is not based on technological performance in isolation, but rather how technology builds knowledge, networks, and capacity. Simply put, plant breeding and natural resource management practices are very "blunt tools for social change" (29); innovation demands sophisticated integration with local partners.

Investment. The growth rate for investments in agricultural R&D declined during the 1990s, particularly for publicly funded agricultural R&D (30), despite research showing that investments in agricultural R&D are one of the most successful ways to alleviate hunger and poverty (31). Developed countries spend, on average, $5.16 on S&T for every $100 of agricultural output, whereas developing countries invest only $0.57 (20).

Continued S&T advancements need to be accompanied by investments in rural infrastructure (physical, market, and finance) and local governance (see table on page 320). Countries lagging behind in these investments simply cannot compete in domestic or international markets. Investments that improve farmers' access to land and water resources are equally vital.

Basic education investments are needed as well. A study of farmers in developing countries showed that those who completed 4 years of elementary education had, on average, 8.7% higher productivity (32).

Agricultural S&T, in and of itself, cannot solve structural inequities and may worsen them by reinforcing existing advantage; nevertheless, S&T can help advance sustainability and development goals with policies and investments that support small-scale sectors. Small farmers in Zimbabwe grew over 90% of the commercial maize crop when markets and services were well organized (33), and Ghanaian cocoa farmers more than doubled their market sales in response to marketing reforms that left them a higher profit share (34). In contrast, an overreliance on free market forces has led to suboptimal investment patterns. For instance, trade arrangements that open national agricultural markets to international competition before basic national institutions and infrastructure are in place can undermine local agricultural sectors (35). The most successful investments will increase the resilience of local and global food systems to environmental and economic shocks.

References and Notes

1. Intergovernmental Panel on Climate Change, *IPCC Fourth Assessment Report: Climate Change 2007* (Cambridge Univ. Press, Cambridge, 2007).
2. Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: Global Assessment Reports* (Island Press, Washington, DC, 2005).
3. D. Molden, Ed., *Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture* (Earthscan, London, 2007).
4. U.N. Environment Programme (UNEP), *Global Environment Outlook (GEO-4)* (UNEP, Nairobi, Kenya, 2007).
5. FAO, *Food Outlook 2007: Global Market Analysis* (FAO, Rome, 2007).
6. R. E. Green, S. J. Cornell, J. P. W. Scharlemann, A. Balmford, *Science* **307**, 550 (2005).
7. FAO, *The State of Food and Agriculture 2007: Paying Farmers for Environmental Services* (FAO, Rome, 2007).
8. *The State of Food Insecurity in the World 2006* (FAO, Rome, 2006).
9. Faostat (agriculture and food statistics), http://faostat.fao.org/site/339/default.aspx.
10. International Assessment of Agricultural Science and Technology for Development (www.agassessment.org).
11. *International Assessment of Agricultural Science and Technology for Development* (Island Press, Washington, DC, 2008).
12. D. Tilman et al., *Science* **292**, 281 (2001).
13. Editor, *Nat. Biotechnol.* **26**, 247 (2008).
14. E. Stokstad, *Science* **319**, 1474 (2008).
15. Editor, *Nature* **451**, 223 (2008).
16. L. T. Evans, *J. Agric. Sci.* **143**, 7 (2005).
17. R. E. Evenson, D. Gollin, *Science* **300**, 758 (2003).
18. D. B. Lobell et al., *Science* **319**, 607 (2008).
19. CGIAR, www.cgiar.org.
20. P. Pardey et al., *Science, Technology, and Skills* (International Science and Technology Practice and Policy (INSTEPP); CGIAR and Department of Applied Economics, University of Minnesota, Rome, 2007).
21. D. P. Delmer, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15739 (2005).
22. C. E. Pray, A. Naseem, *J. Dev. Stud.* **43**, 192 (2007).
23. P. A. Sanchez, *Science* **295**, 2019 (2002).
24. FAO, *Milk and Dairy Products: Post-Harvest Losses and Food Safety in Sub-Saharan Africa and the Near East* (FAO, Rome, 2004).
25. Sustainable Rice Production for Food Security, *Proceedings of the 20th Session of the International Rice Commission*, Bangkok, Thailand, 23 to 26 July 2002 (FAO, Rome, 2003).
26. P. Van Mele, *Agric. Forest Entomol.* **10**, 13 (2008).
27. R. R. B. Leakey et al., *Int. J. Agric. Sustain.* **3**, 1 (2005).
28. M. Dingkuhn et al., *Agric. Water Manage.* **80**, 241 (2006).
29. N. Lilja, J. Dixon, *Exp. Agric.* **44**, 3 (2008).
30. P. Pardey et al., *Agricultural Research: A Growing Global Divide?* (International Food Policy Research Institute (IFRI), Washington, DC, 2006).
31. J. Alston et al., *A Meta-Analysis of Rates of Return to Agricultural R&D: Research Report 113* (IFPRI, Washington, DC, 2002).
32. Sustainable Development Department of the FAO, *Sustainable Rural Development: Progress and Challenges. Education, Training and Extension* (FAO, Rome, 2007).
33. E. S. Nederlof, N. Röling, A. van Huis, *Int. J. Agric. Sustain.* **5**, 247 (2007).
34. T. S. Jane, T. S. Rukuni, in *Zimbabwe's Agricultural Revolution: Managing the Food Economy in the 1990s*, M. Rukuni, C. K. Eicher, Eds. (Univ. of Zimbabwe Publications, Harare, 1994).
35. J. Morrison, A. Sarris, in *WTO Rules for Agriculture Compatible with Development*, J. Morrison, A. Sarris, Eds. (FAO, Rome, 2007).
36. We thank F. Denison, M. Spurlock, M. Burke, and C. Kelder for comments. E.T.K. was supported by an NWO (Netherlands Organization for Scientific Research) fellowship.

MATHEMATICS

Frustration in Complexity

P.-M. Binder

The common thread between all complex systems may not be cooperation but rather the irresolvable coexistence of opposing tendencies.

Defining complexity is frustrating. To some extent, paraphrasing Judge Potter Stewart, "we know it when we see it." With no agreed-upon definition, researchers draw at different places the line that demarcates the limit with its illustrious predecessor, chaos: nonrepeating behavior with sensitive dependence on initial conditions. One common criterion is cooperative behavior: complicated global patterns emerging from local or individual interaction rules between parts of a system. But a more general view of complexity that includes all examples—genetic algorithms, computers, the immune system, the brain, protein folding, the stock market, and systems that evolve and adapt (1–4)—points to the well-defined concept of frustration as a unifying theme.

Consider three spins (which can point up or down) placed at the vertices of a triangle, with the demand that all three be antialigned with each other. This is not possible, and we call such an arrangement "frustrated" (5). The dynamics is also frustrated; any spin flip will fix some of the unsatisfied demands while ruining others. The essence of complexity may lie here: A system without dynamical frustration will either settle to equilibrium or grow without bounds. In contrast, even the simple pendulum is frustrated: The restoring force always opposes the direction of the displacement, producing endless oscillations. Let us look at three manifestations of this phenomenon.

First, the frustration may be geometrical in nature (see the figure, top panel). Consider, for example, mathematical "maps"—equations that take a number or point in space and repeatedly apply a mathematical rule to it, generating a sequence or trajectory. Such maps are among the simplest deterministic rules (dynamical systems) that can produce chaotic behavior (6, 7). Here, chaos is generated by the frustration between the tendencies of close trajectories to converge (folding) or diverge (stretching), giving the system an element of unpredictability. This dynamical frustration combined with determinism gives rise to the beautiful fractal objects such as the Lorenz attractor (see the figure, top panel).

Second, a system can have opposing tendencies at different scales: Cooperative behavior at large scales can oppose local behavior and generate complexity (see the figure, middle panel). Consider, for example, cellular automata, which are simple models of natural systems in which space, time, and the variable describing the local state of the system are all discontinuous. In some cellular automata (8), coherent structures much larger than the smallest units move and collide with other structures, either annihilating each other or giving rise to new structures. Such cellular automata have been used successfully to model fluid and traffic flow. In real life, mismatches between global funding and policies and individual needs in health care and public education have been proposed as the root for the partial failure of both (9). In all the above, the dynamical frustration across scales generates complex behavior. In contrast, systems without distinct coherent structures—such as critical points, sand piles, and forest fire models—are less complex.

Scale frustration may also be introduced through imposed energy or fitness landscapes with peaks or valleys at many scales, such as those generated by spin glasses, protein folding, the traveling salesman problem, and other hard problems. A naïve algorithm looking for an optimal solution cannot cope with this diversity. Inspired by biological evolution, scientists have developed adaptive strategies that operate at multiple scales by allowing replication, mutation, and recombination; these can explore complex landscapes more efficiently.

Genetic algorithms (10), the best-known examples of adaptive strategies, are often considered complex systems themselves.

Third, the frustration can be computational (see the figure, bottom panel). Chomsky (11) proposed a hierarchy of increasingly powerful computing machines, with Turing machines at the top. Certain Turing machines are "universal"—they can do all that modern computers can do, and can thus simulate any of the models described so far. They can thus be considered the most complex of all systems. Computers operate by reading an input, changing their state, and interacting with their own memory. Dynamical frustration increases along the hierarchy as computers spend more effort on memory-related tasks. At its worst, a Turing machine can get caught in an infinite loop and never finish a computation: the famous halting problem.

These three manifestations of dynamical frustration are related. Certain cellular automata and maps are capable of universal computation (12), indicating that even simple dynamical systems can be arbitrarily complex. Multiagent models can generate energy landscapes for their own agents. Chaotic systems can go on forever, but some complex systems better stop: The objective of the immune system is to quickly achieve homeostasis after an external invasion; successful Turing computations halt.

Dynamical frustration is thus a plausible unifying thread for complex systems. Yet, the task of quantifying this concept in a way that includes its three (so far) manifestations is daunting. Fortunately, we have just seen indications that the three are not independent.
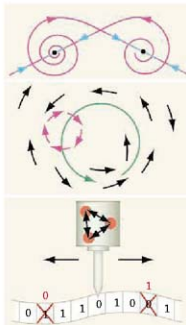
Three manifestations of dynamical frustration. (Top) Geometric frustration. In this schematic of the Lorenz attractor (a butterfly structure in three-dimensional phase space, corresponding to the long-term behavior of a chaotic flow), we can see the interplay between stretching (outward spirals) and folding (arrows directed toward the two fixed points). **(Middle)** Scale frustration. The global motion is generally counterclockwise, but parts of the system may rotate in the clockwise direction. **(Bottom)** Computational frustration. Through complex interactions between the infinite memory tape, the state transition rules of the computing head, the back-and-forth motion (arrows), and the machine rewriting on the memory tape, a Turing machine can perform extremely complex computations—or may never stop computing.

Department of Physics and Astronomy, University of Hawaii at Hilo, Hilo, HI 96720, USA. E-mail: pbinder@hawaii.edu

Moreover, critical systems and the pendulum show us that frustration is not enough. For the latter, nonlinearity and dimensionality play a role. In other cases, especially those involving irregular networks, connectivity can be an important ingredient (13). In stark contrast with the cooperative behavior view, multiple agents are not needed to generate complexity: one Turing machine or one low-dimensional map suffice.

The field of complex systems currently appears as an unfinished mosaic. Many capable researchers are polishing and glu-ing the tiles that may turn it into the queen of all sciences, the science of synthesis and surprise. As we realize how much everything is connected, both cooperation [or emergence (14)] and dynamical frustration can become important tools for our understanding of how the world works.

References

1. R. Badii, A. Politi, *Complexity* (Cambridge Univ. Press, Cambridge, UK, 1997).
2. Y. Bar-Yam, *Introduction to Complex Systems* (Westview, New York, 2003).
3. P. Érdi, *Complexity Explained* (Springer, New York, 2007).
4. *Science*, Special Issue on Complexity, **284** (2 April 1999).
5. G. Toulouse, *Commun. Phys.* **2**, 115 (1977).
6. R. M. May, *Nature* **261**, 459 (1976).
7. S. Strogatz, *Nonlinear Dynamics and Chaos* (Perseus, New York, 2001).
8. S. Wolfram, *Nature* **311**, 419 (1984).
9. Y. Bar-Yam, *Making Things Work* (Knowledge Press, Cambridge, MA, 2005).
10. M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, 1993).
11. N. Chomsky, *Info. and Control* **2**, 137 (1959).
12. P. Koiran, C. Moore, *Theor. Comp. Sci.* **210**, 217 (1999).
13. G. Szabo, G. Fath, *Phys. Rep.* **446**, 97 (2007).
14. N. W. Watkins, M. P. Freeman, *Science* **320**, 323 (2008).

GEOSCIENCE

Natural Complexity

The use of complexity-based approaches may help to advance our understanding of Earth as a system.

Nicholas W. Watkins and Mervyn P. Freeman

Earth is a complex system in which many biological and physical components interact across all space and time scales (1). To understand this system, earth scientists have traditionally built large, multicomponent models. However, it is difficult to know when such a model has become sufficiently detailed for its task and how confident one can be in its predictions. In a generic linear system with feedbacks, Roe and Baker have shown that normally distributed feedbacks give rise to a highly skewed distribution of responses, similar to those seen for climate sensitivity in ensembles of global models (2, 3). Even relatively narrow ranges of uncertainty in the feedbacks can be amplified in the response. Thus, besides refining the feedback uncertainties in traditional earth system models, scientists and policy-makers must explore complementary approaches to modeling.

One such avenue is based on the science of complexity, which describes systems with many strongly interacting parts, concentrating on how the parts connect. A particularly influential strand of complexity science unites two ideas: universality (a given simple property arises in different complex systems) (4, 5) and emergence (complex behaviors arise from simple interaction rules) (6, 7). Emergence is seen, for example, in fish schools (see the figure), which can be modeled using just three rules to show transitions from a stationary swarm via a torus to a parallel group (6). Collective effects like these and others in biology (7) resemble the critical phenomena that

occur when physical systems undergo phase transitions, for example, from solid to liquid. The physics of criticality has been a rich source of insight into emergence (4, 5, 7, 8).

Universality, by contrast, has roots in the 19th century, when the ubiquity of the Gaussian (normal) probability distribution was recognized in many disciplines (4). Today, scientists are increasingly finding that non-Gaussian, in particular power law, probability distributions are also widespread in nature (5, 8), giving rise to rare events with large impacts. Insights have again come from studies of critical phenomena, where power laws are known to be a direct result of the heightened sensitivity to random perturbations conferred by long-range correlations.

Criticality-inspired investigations of nature's complexity have tended to measure universality, typically via power laws, and to then infer the existence of emergence. Closing the loop then requires development of a model, for example, the idea of preferential attachment by which popular Web pages attract proportionally more viewers, giving rise to highly skewed distributions of connections (8). In atmospheric science, long-range correlations in atmospheric temperature records on time scales up to a century have been inferred (9). The existence of such internal long-range correlation would cause variability at very long time scales and thus hamper the detection of externally forced climate change. In ecology, distributions in the durations of, or times or distances between, feeding of numerous species (including bumblebees and deer) have been reported to be highly skewed. This was attributed to optimal foraging strategies based on non-Gaussian random walks (10).

Complexity explained. The behavior of fish schools and other natural systems can be understood through the concepts of emergence and universality.

However, many different mechanisms can give rise to highly skewed distributions (5). It can also be very difficult to distinguish between true long-range correlation and shorter-range mechanisms (11). This is a generic problem with measured fractals, which can typically only be observed over a narrow spatial range (12). Similar problems arise in ecology, where the inference of highly skewed distributions in foraging has been shown to be problematic when different candidate distributions are compared and the restricted observational range is allowed for (10).

Despite these practical problems, the rapid recent developments in the theory of complex networks—that is, sets of interacting, linked

British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK. E-mail: nww@ bas.ac.uk; mpf@bas.ac.uk

nodes in subject areas as diverse as mobile communications (cell phones), ecology (food webs), genomics, and the Internet—are indicative of how progress can be made. A key ingredient has been the willingness to modify initially proposed mechanisms in the light of rich new data sets. For example, recent work has demonstrated that the stability of food webs depends not on network complexity (or geometry) as such but on the overall patterning of link strengths (13). This work is the culmination of 20 years of theoretical and experimental work that increasingly took into account dynamic factors and the energy flow in food webs.

To fully realize the potential of complexity science in environmental science, much more work is needed on distinguishing the complex (emergent) from the just complicated in the presence of many possible explanatory mod-els and imperfect data (5, 14). This will be aided by further development of other rigorous, objective measures (14, 15) in addition to those inspired by critical phenomena. It may even be possible to quantify the onset of increasing complexity based on an analogy between criticality and fluid turbulence (16). If this analogy can also be applied successfully to an ecological model, it would be a dramatic example of how critical phenomena can inform earth system science.

References and Notes

1. V. Smil, *The Earth's Biosphere: Evolution, Dynamics and Change* (MIT Press, Cambridge, 2002).
2. G. H. Roe, M. B. Baker, *Science* **318**, 629 (2007).
3. M. R. Allen, D. J. Frame, *Science* **318**, 582 (2007).
4. P. Ball, *Critical Mass: How One Thing Leads to Another* (Arrow Books, London, UK, 2004).
5. D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* (Springer, Berlin, ed. 2, 2004).
6. I. D. Couzin et al., *J. Theor. Biol.* **218**, 1 (2002).
7. D. T. Suzuki, *Philos. Trans. R. Soc. B* **361**, 5 (2006).
8. G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Univ. Press, Oxford, UK, 2007).
9. A. Bunde et al., in *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes*, A. Bunde, J. Kropp, H. J. Schellnhuber, Eds. (Springer, Berlin, Germany, 2002), pp. 170–191.
10. A. M. Edwards et al., *Nature* **449**, 1044 (2007).
11. D. Maraun et al., *Nonlinear Proc. Geophys.* **11**, 495 (2004).
12. D. Amiri et al., *Science* **279**, 39 (1998).
13. A.-M. Neutel et al., *Nature* **449**, 599 (2007).
14. C. Shalizi, in *Complex Systems Science in Biomedicine*, T. S. Deisboeck, J. Y. Kresh, Eds. (Springer, New York, 2006), pp. 33–114.
15. P.-M. Binder, *Science* **320**, 322 (2008).
16. S. C. Chapman et al., http://arxiv.org/abs/0707.3958.
17. The authors acknowledge conversations with colleagues at BAS and with participants at the meeting on Natural Complexity: Data and Theory in Dialogue at the Law Faculty and Clare College in Cambridge, UK, from 13 to 17 August 2007.

Graphene Nanoelectronics

Graphene quantum dots offer a new approach to quantum nanoelectronics.

R. M. Westervelt

Semiconductor technology has taken us a long way by making devices of ever smaller size. But eventually, as the transistors approach the size of molecules, quantum effects become important. What will then be the form of future nanoelectronic devices? Can quantum mechanics be used to control device operation? And can they operate at reasonable temperatures? Nanoscale transistors made from graphene may provide ways to address these questions. On page 356 of this issue, Ponomarenko et al. (1) describe graphene single-electron transistors etched to sizes as small as ~30 nm, which have quantum-confined energy states, and control the motion of single electrons (see the figure). This complements investigations of single-electron transistors from graphene flakes (2), quantum interference devices (3), and ~200-nm etched graphene dots (4).

Graphene is an unusually simple material with startling new properties (5). It consists of a single atomic layer of carbon atoms. One can prepare a graphene sample by pulling an atomic layer off of a graphite crystal with a piece of
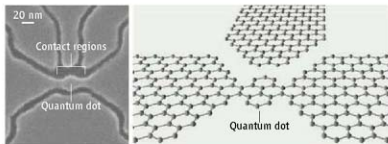
Graphene devices. (Left) Scanning electron micrograph of a graphene-based single-electron transistor. The dark areas are gaps in the photoresist mask where graphene is removed by plasma etching. The small quantum dot in the center is connected to contact regions through narrow constrictions, and there are four side gates. **(Right)** Schematic showing a hypothetical graphene-based single-electron-transistor based on a quantum dot that contains only a few benzene rings.

sticky tape, and then pressing the tape down onto a silicon substrate. After pulling the tape free, small islands of graphene remain, recognizable with an optical microscope. Graphene layers are tough—the carbon atoms bind together in a hexagonal lattice—and can be freely suspended over a trench in the substrate.

It is the behavior of electrons in single graphene layers (5–9) that is opening the way for new kinds of devices. In a typical semiconductor, electrons and holes (the positively charged carriers) occupy the conduction and valence energy bands, respectively. A finite energy called the energy gap must be added to a particle to move it from the valence to the conductance band. Electrons and holes move like regular particles: They have a mass, their speed increases from zero as they are accelerated, and their kinetic energy is proportional to the square of their speed. In graphene, the behavior of electrons and holes is very different: The particles move with a constant speed v_g that does not depend on their kinetic energy E. This is similar to the behavior of photons, which always travel at the speed of light c. In graphene, the speed of electrons and holes is slower than light by a factor of 300.

Graphene also differs because there is no energy gap: The conduction and valence bands are shaped like an inverted pair of cones that meet in a single point at $E = 0$ in momentum space. The relativistic form of the energy bands is new for solids, and it changes the rules that govern how electrons move through a graphene-based device.

Novoselov et al. (8) found that one could control the carrier density in graphene in an unusual way. A positive gate voltage attracts electrons, so the Fermi energy lies at positive energies $E > 0$ in the cone-shaped conduction band. As the gate voltage is made less positive, the density of electrons decreases, but the conductance G does not fall to zero at $E = 0$, as it would in a normal semiconductor. Instead, it reaches a minimum value comparable to the

School of Engineering and Applied Sciences, and Department of Physics, Harvard University, Cambridge, MA 02138, USA. E-mail: westervelt@seas.harvard.edu

quantum of conductance $G_0 = e^2/\hbar$, where e is the electron charge, and $\hbar$ is Planck's constant. As the gate voltage is moved to negative values, the conductance increases again, because holes are attracted. Because there is no energy gap, it is not possible to deplete the carriers completely and drive the conductance to zero. New ways must then be found to make transistors.

An even more unusual phenomenon occurs in graphene: The Klein paradox allows a relativistic particle to pass freely through a tall barrier of great width (10), whereas an ordinary particle would bounce backward, like a baseball after it hits a wall. As an electron approaches a potential barrier created by a gate, or by charged traps in the substrate, the electron's kinetic energy is reduced to zero as it hits the barrier. But instead of backscattering the electron, as one would expect, the barrier converts the electron into a hole, which is then attracted to the barrier potential and moves forward freely. When the hole exits the far side of the barrier, it turns back into an electron. The Klein paradox allows electrons to move through a graphene layer as if it were ideal. Graphene samples are not perfect, but are divided into pools of electrons or holes. Despite these imperfections, the charge carriers can move ~ 0.3 μm at room temperature with little scattering (5). This allows one to make ballistic transistors from a simply prepared material.

Electrons in graphene show strong quantum behavior, even at room temperature, in part becaue of their freedom of motion. In the quantum Hall effect, the Hall conductance is quantized in integer multiples of G_0. To observe this effect, high quality samples, strong magnetic fields, and low temperatures ~4 K are usually needed. The quantum Hall effect has been observed in graphene (11, 12)—the quality of the data is very high, despite the simple methods used to produce the samples. It is interesting that the quantization of Hall conductance steps is different for graphene: They are spaced by $4G_0$, with the lowest steps at $\pm 2G_0$. These changes are caused by graphene's unusual band structure in which the electrons and holes travel at a constant speed. Recently, the quantum Hall effect in graphene has been observed at room temperature (13), which demonstrates its potential for quantum devices.

The freedom of motion associated with the Klein paradox creates a problem: How does one confine charge carriers inside a device? A simple approach is to cut the graphene layer into the right shape, as for the quantum dot in the figure. Quantum confinement of electrons can also be used to control their motion. For example, a narrow ribbon of graphene (14) can effectively create an energy gap between

the electron and hole bands, the magnitude of the gap being inversely proportional to the ribbon width. With an energy gap, one can then deplete the carrier concentrations with a barrier or gate, as with a conventional device.

A quantum dot created by etching a graphene layer is shown on the left in the figure. A pool of electrons is confined inside a graphene disc that is connected to its leads by two narrow constrictions; a graphene side gate is used to tune the dot charge. Because they are very narrow, the conductance of these constrictions can be reduced to values $<G_0$, which allows the dot to trap individual electrons at low temperatures. In this regime, the quantum dot acts as a single-electron transistor, and its conductance shows a periodic set of peaks as the gate voltage is increased. By making very small graphene dots with sizes <100 nm, Pomonarenko et al. (1) were able to enter a regime where the conductance peaks are no longer periodic, but are instead controlled by the energy of the individual quantum states of electrons trapped inside. The measured energy distribution showed the electrons behaved as a chaotic quantum system (7), as one might expect for a dot that is not perfectly round. By etching the dots even smaller, they were able to achieve transistor operation at room temperature.

Graphene is an exciting new material with unusual properties that are promising for nanoelectronics. The carriers move freely, ignoring barriers created by imperfections, and they show quantum effects at room temperature. Through advances in fabrication and characterization building on those of Ponomarenko et al. (1), it may become possible to make quantum dots so small that they approach the molecular scale (see figure, right panel). The future should be very interesting.

References

1. L. A. Ponomarenko et al., Science 320, 356 (2008).
2. J. S. Bunch, Y. Yaish, M. Brink, K. Bolotin, P. L. McEuen, Nano Lett. 5, 287 (2005).
3. F. Miao et al., Science 317, 1530 (2007).
4. C. Stampfer et al., Appl. Phys. Lett. 92, 012102 (2008).
5. A. K. Geim, K. S. Novoselov, Nat. Mater. 6, 183 (2007).
6. P. R. Wallace, Phys. Rev. 71, 622 (1947).
7. M. V. Berry, R. J. Mondragon, Proc. R. Soc. London Ser. A 412, 53 (1987).
8. K. S. Novoselov et al. Science 306, 666 (2004).
9. K. S. Novoselov, Proc. Natl. Acad. Sci. U.S.A. 102, 10451 (2005).
10. M. I. Katsnelson, K. S. Novoselov, A. K. Geim, Nat. Phys. 2, 620 (2006).
11. K. S. Novoselov et al., Nature 438, 197 (2005).
12. Y. Zhang, J. W. Tan, H. L. Stormer, P. Kim, Nature 438, 201 (2005).
13. K. S. Novoselov et al., Science 315, 1379 (2007).
14. M. Y. Han, B. Ozyilmaz, Y. Zhang, P. Kim, Phys. Rev. Lett. 98, 206805 (2007).

MEDICINE

The Ultimate Model Organism

Atul J. Butte

A deeper understanding of disease requires a database of human traits and disease states that is integrated with molecular information.

This month, the scientific community celebrates the 25th anniversary of GenBank, the open access database of DNA sequences and the molecules they encode. Heralded as one of the earliest bioinformatics community projects, it has fueled our need to understand how this information can be linked to physiology and disease. Since then, biocomputational, informatics, and statistical methods have been used to relate sequences and molecules to diseases. But as highlighted in meetings such as last month's Summit on Translational Bioinformatics (1), the same high-bandwidth measurement style

that has accelerated the molecular and genetic study of disease must be practiced in physiology if we are to gain a deeper understanding of normal and impaired health.

Within the last 5 years, systematic studies on the commonalities (2) and differences (3) across diseases have shown that particular biological signaling pathways and modules share similar properties. Other studies have shown that diseases that resemble each other can share genes with variants (4, 5) or share genes coding for proteins that interact with each other (6). So many diseases have now been studied that publicly available data can be used to find genes with common changes in expression for each condition (7).

The difficulty with interpreting such analyses lies with how diseases are defined. The definition of a disease is often specified

Stanford Center for Biomedical Informatics Research, Department of Medicine and Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA, and Lucile Packard Children's Hospital, Palo Alto, CA 94304, USA. E-mail: abutte@stanford.edu

by a particular knowledge base and is thus subject to limitations and biases. For example, a network built from a knowledge base of monogenic diseases (those associated with a single gene) may not be generalizable to more common diseases caused by multiple genes (5). More recently described diseases, such as sudden infant death syndrome, may be less well characterized, so searches for gene variants by matching syndromes through clinical descriptions could yield false-negative predictions (6). In some studies, gene expression or genotyping samples are studied based on clinical disease labeling by physicians, and thus, genes found to be "associated" with a condition do not yet fully explain the observed traits present in a disease. Moreover, there is a growing movement toward direct-to-consumer testing, with the promise that consumer-provided DNA samples could be used for genome-wide association studies. It remains to be seen how such samples can be analyzed when the "assignment" of phenotype and disease is provided by a consumer.

On the other hand, parallel measurements of physiological variables have been successfully linked to genetic markers in animal models of diseases, such as hypertension (8). These efforts have driven the organization of efforts such as the Physiome Project (9), an international collaboration to model the human body through computational methods that integrate biochemical, biophysical, and anatomical information about cells, tissues, and organs. There have also been calls for a Human Phenome Project (10), whose goal is to establish databases of phenotypes that are associated with physiology, to determine their relation with genes and proteins. Data for some complex human physiological traits are already publicly available for analysis at resources such as PhysioNet (11).

Yet, the current approach for defining phenotypes for molecular discovery is not adequate, and therefore, doesn't optimize the use of physiological data. Phenotypes—traits ranging from height and weight to glucose metabolism, predisposition to disease, and behavioral characteristics—and their differences between individuals can be due to environmental influences and/or genetic variation. One solution for defining richer phenotypes is to take advantage of clinical measurements, which are born from physiological measurements. Enormous numbers of clinical tests are performed each year, and are increasingly being captured in electronic health records, along with patient interventions (medications or procedures). These kinds of data could be used to answer basic biological questions (12, 13). Mathematical arrays of such data have already been assembled from hospital-based clinical measurements or epidemiological information and have successfully identified biomarkers for human maturation and aging (14, 15). Connections between clinical findings and molecular measurements can also now be tested across a large set of findings and molecules. For example, gene expression profiles of individual liver cancer samples have been predicted by prior radiological findings on abdominal computed tomography scans (16).

How can we take advantage of the petabytes of clinical measurements on patients for whom genetic or genomic measurements may not yet have been obtained? The same broad consideration across diseases used successfully in molecular studies could also be applied to clinical measurements. For example, suppose three diseases are separately considered by a quantitative clinical laboratory test measurement (obtained from an electronic health record), and a gene

Information intersection. Three diseases may be separately considered by a quantitative clinical laboratory test measurement and a gene expression measurement (from a public repository of gene expression). Associations can be discovered between molecular and clinical measurements, even when these measurements are not made using the same samples or patients. For example, Disease A, when studied across all patients and time points, shows a high average level of a clinical test (red line), and a low level of a gene (blue line). The distribution of gene and clinical measurements are shown by sampling from both independent data sets (colored regions). The trend across the three diseases shown is that as a disease shows less of a clinical measurement in patients, it shows more expression of a particular gene.

expression measurement (from a public repository of gene expression data) (see the figure).Within a disease, the distribution of gene and clinical measurements can be shown, but whether the gene and clinical measurements correlate cannot be determined, as the measurements are not taken from the same patients. But trends might be observed across the three diseases. For instance, as a disease shows more or less of a clinical measurement in patients, microarray samples of the disease may show more or less expression of a particular gene. Thus, associations could be discovered between molecular and clinical measurements, even when these measurements are not made using the same samples or patients. Instead of studying samples or patients as data points in the traditional reductionist manner, one could study and plot diseases.

But there are challenges to using clinical data as physiological measurements. Access issues to patients' private health information can dissuade basic researchers from using clinical measurements. Although patient data could be deidentified and patients approached for informed consent, much clinical data exists as documents that are difficult to deidentify and/or sift through using automated processes. Even as these challenges are addressed, purely numerical quantitative clinicians measurements could be used to start, as these are the easiest to deidentify and analyze.

Whereas there are public international repositories for many molecular measurements, we do not yet have an equivalent for deidentified clinical measurements. There are multiple reasons for this. The fear that personal medical information could be inappropriately released is a powerful disincentive for sharing. Clinical data may also be viewed by clinicians and hospitals as a "trade secret," and only recently are data on performance and quality being published. This fear could be averted if health care networks pooled deidentified data sets, thus deidentifying the source of care as well. Clinical researchers are also justifiably protective of the resources they create and might fear missing a discovery within their own patient cohort. Availability agreements could address retention of rights, intellectual property, and publication embargoes. Instead of viewing data availability as a disadvantage, clinical researchers and institutions should be encouraged to look at the success of resources such as GenBank to see how the public availability of deidentified data can yield many more discoveries when shared.

A population of well-supported and trained scientists and physicians must be nurtured to relate the enormity of physiologi-

cal and clinical measurements to molecular measurements. The multiscale models of health they build will finally yield an understanding of disease that is more than just the sum of its parts.

References

1. www.amia.org/meetings/stb08/
2. D. R. Rhodes et al., Proc. Natl. Acad. Sci. U.S.A. **101**, 9309 (2004).
3. E. Segal, N. Friedman, D. Koller, A. Regev, Nat. Genet.

36, 1090 (2004).
4. The Wellcome Trust Case Control Consortium, Nature **447**, 661 (2007).
5. K. I. Goh et al., Proc. Natl. Acad. Sci. U.S.A. **104**, 8685 (2007).
6. K. Lage et al., Nat. Biotechnol. **25**, 309 (2007).
7. A. J. Butte, I. S. Kohane, Nat. Biotechnol. **24**, 55 (2006).
8. M. Stoll et al., Science **294**, 1723 (2001).
9. P. J. Hunter, T. K. Borg, Nat. Rev. Mol. Cell Biol. **4**, 237 (2003).
10. N. Freimer, C. Sabatti, Nat. Genet. **34**, 15 (2003).
11. http://www.physionet.org/

12. N. S. Sung et al., JAMA **289**, 1278 (2003).
13. P. R. Payne et al., J. Investig. Med. **53**, 192 (2005).
14. D. P. Chen et al., Pac. Symp. Biocomput. 2008, 243 (2008).
15. A. Fliss, M. Ragolsky, E. Rubin, in Summit on Translational Bioinformatics Proceedings, San Francisco, CA, 10 to 12 March 2008 (AMIA, Bethesda, MD, ISCB, La Jolla, CA, 2008), p. 11.
16. E. Segal et al., Nat. Biotechnol. **25**, 675 (2007).

CELL SIGNALING

Wnt Moves Beyond the Canon

Bruce Bowerman

A secreted protein associated with the metastasis of a melanoma harnesses cell adhesion proteins and the cytoskeleton to direct cell motility.

The cellular signaling pathways that respond to Wnts, a highly conserved family of secreted proteins, control numerous and different animal developmental processes, particularly those that govern cell fate and patterning. Often, components of Wnt signaling pathways malfunction and foster deadly cancers (1). Although intensely studied, much about Wnt signaling remains enigmatic. Whereas most signaling events that target gene expression in response to Wnt are referred to as canonical pathways (the more commonly observed sequence of molecular events), noncanonical Wnt signaling often targets the cytoskeleton. On page 365 of this issue, Witze et al. (2) advance our understanding of a noncanonical Wnt signaling pathway (Wnt5a) that promotes the invasive behavior of melanoma cancer cells. Their findings advance our understanding of cancer metastasis and highlight the diversity of mechanisms through which Wnt signals influence cells.

The canonical Wnt signaling pathway was deduced from genetic studies in the fruit fly Drosophila melanogaster and is widely conserved in other animals (1). Wnt binds to Frizzled, its transmembrane receptor, prompting cytosolic Dishevelled proteins to prevent the otherwise constitutive proteolytic destruction of another cytosolic protein, β-catenin. Stabilized β-catenin then transits to the nucleus, where it converts transcriptional repressors called T cell factors into activators, driving the expression of target genes that specify cell fate. By contrast, one cannot so easily summarize noncanonical Wnt signaling. The best-understood example was discovered in Drosophila, where it polarizes epithelial cells along a common axis (3, 4). This pla-

nar cell polarity requires the canonical Wnt signaling components Frizzled and Dishevelled as well as proteins that do not participate in the canonical pathway. Although related signaling pathways that respond to Wnt influence cell migration during early vertebrate embryogenesis (gastrulation), the full extent of their conservation remains unclear. More recently, genetic studies in the nematode Caenorhabditis elegans have identified noncanonical Wnt signaling pathways that bear little or no resemblance to the one that controls planar cell polarity (5–9). The only unifying theme is that all of these examples somehow differ from the canon.
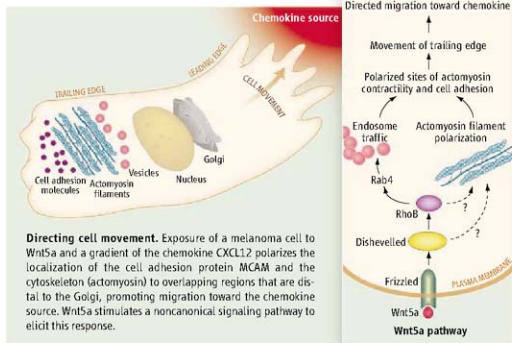
Nevertheless, a shared feature of many noncanonical Wnt pathways is their targeting of the cytoskeleton, with important implications for cancer. For example, expression of noncanonical Wnt5a correlates with metasta-

tic melanoma invasiveness in humans, and exposure to Wnt5a, but not other Wnts, promotes invasiveness in melanoma cell lines (10). Witze et al. investigated how Wnt5a polarizes the cytoskeleton to promote directional motility in cultured melanoma cells. They show that this response is permissive, requiring an independent directional cue (a chemokine). Dispersed cells within a chemokine gradient require exposure to Wnt5a for polarization and migration, but exposure to Wnt5a alone has little effect.

To understand this noncanonical and permissive response, Witze et al. used fixed- and live-cell imaging to examine the localization and movement of proteins within single cells shortly after exposure to Wnt5a. They examined several proteins, including a melanoma cell adhesion molecule previously implicated in metastatic melanoma (11), a Frizzled

Directing cell movement. Exposure of a melanoma cell to Wnt5a and a gradient of the chemokine CXCL12 polarizes the localization of the cell adhesion protein MCAM and the cytoskeleton (actomyosin) to overlapping regions that are distal to the Golgi, promoting migration toward the chemokine source. Wnt5a stimulates a noncanonical signaling pathway to elicit this response.

Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA. E-mail: bbowerman@molbio.uoregon.edu

receptor, and cytoskeletal components (actin microfilaments and a nonmuscle myosin, together called actomyosin). In response to Wnt5a, all of these proteins become polarized in overlapping bands within the trailing edge of migrating cells (see the figure). Polarized bands of ac-tomyosin contract to lift the trailing edge and, together with cell process extensions at the leading edge, promote migration toward a chemokine source. Multiple pulses of these transient polarizations occur within single cells over time.

In exploring the signaling pathway that mediates this Wnt5a response, Witze *et al.* discovered that actin microfilaments and the cell adhesion protein depend on each other for polarized localization and that both are required for myosin polarization. Like some other noncanonical Wnt pathways, the cytosolic proteins Dishevelled and Protein Kinase C are required. But in contrast to noncanonical pathways that influence actin through the small guanosine triphosphatase (GTPase) RhoA, Wnt5a acts through RhoB, which controls the movement of intracellular vesicles called endosomes. Furthermore, the authors observed an enrichment of multivesicular bodies at the polarized cell edge, and polarization required both the GTPase Rab4 (known to regulate multivesicular body formation) and dynamin (known to control the formation of endosomes at the cell surface). Witze *et al.* also found that polarization of the cell adhesion protein occurs by its transit through the Golgi. The authors conclude that Wnt5a polarizes cells by promoting the recycling of membrane

components to specific surface sites. The concurrent or subsequent recruitment of actomyosin and other factors to these sites defines the trailing edge of the cell and thus the direction of migration.

This admirably integrated view of both cell surface and actomyosin polarization raises important questions. For instance, it's not clear how polarization of the membrane and cytoplasm are coupled, or how different proteins are targeted to the same sites. Also, how does Wnt5a signaling interface with the graded chemokine to ensure that polarization occurs at cell surfaces distal to the chemokine source? Intriguingly, Witze *et al.* observed that some responding melanoma cells polarized in the opposite (proximal) direction. Both distal and proximal polarizations require Wnt5a, with a preponderance of distal events promoting migration toward the chemokine source. Thus, the chemokine and Wnt5a are more tightly coupled to the same axis than they are to the same direction along that axis.

It is interesting to compare Wnt5a to other noncanonical Wnt signals that polarize single cells. One common theme is a requirement for actin. Actin microfilaments and myosin are required for planar cell polarity in *Drosophila* (*3*), and microfilaments are required in four-cell-stage *C. elegans* embryos for the Wnt-mediated induction of an asymmetric cell division that segregates endoderm from mesoderm (*6*). However, planar cell polarity acts through RhoA (*8*), whereas Wnt5a acts through RhoB. Furthermore, the *C. elegans*

Wnt signal that segregates endoderm from mesoderm acts instructively and requires a permissive signal (a tyrosine kinase) (*12*), in contrast to the permissive action of Wnt5a in the presence of an instructive chemokine signal. Once again, noncanonical Wnt pathways present a bewildering mix of similarities and differences. Appropriately, the Wnt field is evolving toward the use of individual pathway names based on key molecules and properties, and away from the historical distinction between canonical and noncanonical. Presumably, as we learn to speak more clearly about different Wnt pathways, we also will come to better understand their different molecular mechanisms of action.

References and Notes

1. C. Y. Logan, R. Nusse, *Annu. Rev. Cell Dev. Biol.* **20**, 781 (2004).
2. E. S. Witze, E. S. Litman, G. M. Argast, R. T. Moon, N. G. Ahn, *Science* **320**, 365 (2008).
3. M. T. Veeman, J. D. Axelrod, R. T. Moon, *Dev. Cell* **5**, 367 (2003).
4. P. N. Adler, *Dev. Cell* **2**, 525 (2002).
5. F. D. Park, J. R. Tenlen, J. R. Priess, *Curr. Biol.* **14**, 2252 (2004).
6. C. J. Thorpe, A. Schlesinger, B. Bowerman, *Trends Cell Biol.* **10**, 10 (2000).
7. M. A. Herman, M. Wu, *Front. Biosci.* **9**, 1530 (2004).
8. K. Mizumoto, H. Sawa, *Trends Cell Biol.* **17**, 465 (2007).
9. T. Kaletta, H. Schnabel, R. Schnabel, *Nature* **390**, 294 (1997).
10. A. T. Weeraratna *et al.*, *Cancer Cell* **1**, 279 (2002).
11. K. Satyamoorthy, J. Muyrers, F. Meier, D. Patel, M. Herlyn, *Oncogene* **20**, 4676 (2001).
12. B. Goldstein, H. Takeshita, K. Mizumoto, H. Sawa, *Dev. Cell* **10**, 391 (2006).
13. Work in the author's laboratory is funded by grants from NIH.

MUSIC THEORY

Geometrical Music Theory

Rachel Wells Hall

Musical operations, such as transpositions, can be expressed as symmetries of *n*-dimensional space.

Music theorists normally use discrete mathematics, such as set and finite group theory, to describe musical relations; they also invoke geometry in modeling musical objects such as chords, rhythms, and scales. However, no unified geometric perspective has hitherto emerged. On page 346 of this issue, Callender *et al.* (*1*) demonstrate that many musical terms can be understood as expressing symmetries of *n*-dimensional space, where each dimension represents a voice in the score. Identifying—gluing together—points related by these symmetries

produces exotic musical spaces (orbifolds) that subsume a large number of geometric models previously proposed.

The use of mathematics to describe, analyze, and create music goes back millennia. Mathematical questions have previously emerged in music theory that are appealing, nontrivial, and, in several cases, connected to other scientific fields, such as physics, dynamical systems, and crystallography (*2–4*). Math inspired composers such as Schoenberg, Messiaen, and Xenakis; musical investigations have even motivated mathematical discoveries (*5*).

However, the geometrical music theory proposed by Callender *et al.* stands out both for the breadth of its musical implica-

tions and the depth of its mathematical content. It suggests a wealth of new techniques for studying musical theoretical topics, including chord similarity and melodic contour, and provides attractive visualizations of harmonic relationships. Although the authors' use of contemporary geometry departs from the discrete mathematics normally used to describe music, they make a convincing argument for the validity of their models.

Although most musicians are not mathematicians, both groups think in degrees of abstraction. Music theorists invoke the mathematical concept of equivalence class—a set of objects that are "the same" if we ignore certain information—to describe these musical struc-

Department of Mathematics and Computer Science, Saint Joseph's University, Philadelphia, PA 19131, USA. E-mail: rhall@sju.edu

tures. For example, middle C is a particular pitch (frequency measured on a logarithmic scale), and the letter name (or pitch class) "C" refers to any pitch that is a whole number of octaves away from it. Any collection of C, E, and G notes is a C major chord.

Callender et al. (1) show that many musical terms describe equivalence classes under combinations of five basic OPTIC relations: octave shifts, permutation (reordering), transposition (the relation between pitches sharing the same succession of intervals, regardless of their starting note), inversion (turning a sequence upside down), and cardinality equivalence (ignoring repetitions). The authors go beyond traditional music theory by showing that there are two distinct ways in which these symmetries can apply to chord progressions: uniformly, where the same symmetry applies to each chord, and individually, where distinct symmetry operations apply to the harmonies in a progression. This distinction allows Callender et al. to formalize the relations among a large number of musical terms as in table S1 of their paper (1).

Following Tymoczko's previous work (6), they develop geometrical models of these relations. Pitches correspond to real numbers; each point in Euclidean n-dimensional space (that is, a space with no underlying curvature) represents a sequence of n pitches. Points near each other differ by microtones (for example, a C major chord played out of tune lies near an in-tune version) and are in this sense similar.

With the exception of cardinality equivalence, each of the OPTIC relations corresponds to a symmetry operation. For example, permuting the first two entries in a sequence is equivalent to reflection in a plane of n − 1 dimensions (a hyperplane), and transposing a sequence is equivalent to a translation. Gluing together points in n-dimensional space on the basis of their memberships in equivalence classes creates families of quotient spaces, or orbifolds. Many of these mathematical spaces are geometrically and
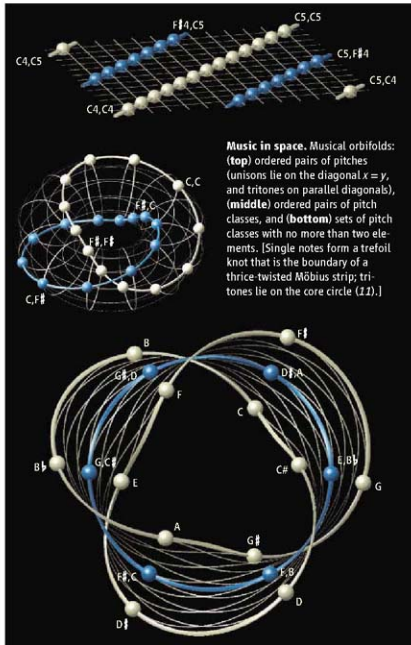
Music in space. Musical orbifolds: **(top)** ordered pairs of pitches (unisons lie on the diagonal $x = y$, and tritones on parallel diagonals), **(middle)** ordered pairs of pitch classes, and **(bottom)** sets of pitch classes with no more than two elements. [Single notes form a trefoil knot that is the boundary of a thrice-twisted Möbius strip; tritones lie on the core circle (11).]

topologically complex; they may be non-Euclidean, meaning that there may be more than one line connecting two points (see the figure). These features have musical ramifications. Callender et al. provide a comprehensive catalog of their geometrical spaces in table S2 (1).

Distance and movement within these spaces have natural meanings. A musical score represents pitch vertically and time horizontally. Each vertical harmonic state (the notes played at one time) corresponds to a pitch sequence, ordered by instrumental voice; a score for n voices becomes a succession of points in n-dimensional space, where each coordinate tracks the melody sounded by an individual voice. Voice leadings—mappings between adjacent chords in a score—correspond to directed line segments connecting harmonic states. The size of a voice leading

depends on the amount of vocal movement between states. Composers typically try to minimize this movement—a problem equivalent to finding the shortest path between two points.

Although the design of efficient voice leadings is central to traditional compositional practice, the systematic study of voice leading is comparatively recent (7–10). Geometrical music theory essentially represents a voice leading by a vector that connects a source chord to a target chord. The authors show the musical distinction between individual and uniform applications of symmetries corresponds to the geometrical notion of a bundle, which allows them to define analogs of voice leadings between abstract musical objects such as major chords and minor chords.

Geometrical music theory suggests new directions for research in traditional music theoretical topics, such as chord and voice leading similarity, and new tools for teaching and conceptualizing music, and will perhaps inspire composers. Moreover, one can envision practical applications of geometrical music theory, such as in the design of music visualization tools, interactive musical toys, or even new musical instruments.

References

1. C. Callender, I. Quinn, D. Tymoczko, Science 320, 346 (2008).
2. J. Clarke, R. Voss, Nature 258, 317 (1975).
3. J. Douthett, R. Krantz, J. Math. Phys. 37, 7 3334 (1996).
4. G. Toussaint, in Discrete and Computational Geometry Japanese Conference, JCDCG 2004, Tokyo, Japan, 8 to 11 October 2004, Revised Selected Papers Proceedings of the Japan Conference on Discrete and Computational Geometry (Springer, Berlin, 2005).
5. D. Knuth, The Art of Computer Programming, vol. 4, fascile 4, Generating All Trees: History of Combinatorial Generation (Addison-Wesley, Upper Saddle River, NJ, 2006).
6. D. Tymoczko, Science 313, 72 (2006).
7. J. Roeder, thesis, Yale University (1984).
8. R. Morris, Mus. Theory Spectrum 20, 2 175 (1998).
9. J. Straus, Mus. Theory Spectrum 25, 2 305 (2003).
10. R. Cohn, Mus. Theory Online 9, 4 (2003).
11. C. Tuffley, Algebr. Geom. Topol. 2, 1119 (2002).

Qinghaosu (Artemisinin): The Price of Success

N. J. White

Artemisinin and its derivatives have become essential components of antimalarial treatment. These plant-derived peroxides are unique among antimalarial drugs in killing the young intraerythrocytic malaria parasites, thereby preventing their development to more pathological mature stages. This results in rapid clinical and parasitological responses to treatment and life-saving benefit in severe malaria. Artemisinin combination treatments (ACTs) are now first-line drugs for uncomplicated falciparum malaria, but access to ACTs is still limited in most malaria-endemic countries. Improved agricultural practices, selection of high-yielding hybrids, microbial production, and the development of synthetic peroxides will lower prices. A global subsidy would make these drugs more affordable and available. ACTs are central to current malaria elimination initiatives, but there are concerns that tolerance to artemisinins may be emerging in Cambodia.

In the fourth decade of the 17 century, Jesuits brought the bark of a Peruvian tree (later named Cinchona) to Europe. This provided for the first time a specific remedy for agues, periodic fevers that were prevalent throughout much of the continent, particularly in and around marshy areas (mal-aria or bad air). In 1880, Alphonse Laveran identified the intra-erythrocytic protozoan parasite that caused malaria. The Cinchona alkaloids (quinine, quinidine, cinchonidine, and cinchonine) were shown to arrest the development of the malignant tertian malaria parasites (*Plasmodium falciparum*) at the mature trophozoite stage (after the first third of their 2- day intraerythrocytic life cycle) and thereby prevent their multiplication in the red blood cells (Fig. 1). Today, the Cinchona alkaloids are giving way to the products of a ubiquitous annual shrub (*Artemesia annua*, or huang hua hao, but often called qinghao) (Fig. 2).

The antimalarial properties of the traditional Chinese medicine qinghaosu (artemisinin) were discovered by Chinese scientists in 1971 who performed low temperature ethyl ether extractions of *Artemesia annua*. In a research effort, apparently prompted by the requests of Ho Chi Minh to Zhou En Lai for antimalarial drugs to protect his Vietnamese troops, the scientists identified the active antimalarial principle, characterized its physicochemical properties, conducted in vitro, animal, and human studies, and synthesized derivatives of the more potent dihydroartemisinin (DHA). Artemisinin was first announced to the rest of the world in 1979 (1). At first, biological chemists were puzzled by the apparent stability of the hitherto unknown 15-carbon (sesquiterpene) peroxide structure. A full chemical synthesis was reported 4 years later (2), although, as for quinine, this remains too expensive for commercialization. Trials reporting efficacy in both uncomplicated and severe malaria soon followed (3, 4), but progress thereafter was slow. Instead of accepting the compounds the Chinese had developed, the World Health Organization (WHO)–Special Program for Research and Training in Tropical Diseases (TDR), the pharmaceutical industry, and the U.S. Army elected to develop their own compounds. Time and money was wasted developing artemotil (arteether), the ethyl ether of DHA, an oil-based formulation for intramuscular injection, which the Chinese scientists had discarded earlier in favor of the almost identical artemether (the methyl ether of DHA) (Fig. 3). Initially, the orally active compounds and the water-soluble artesunate,

which could be given intravenously, were neglected entirely outside China, but with worsening resistance to all available antimalarials in Southeast Asia, researchers there began to investigate the compounds from China, and an increasing evidence base accumulated supporting the rapidity, reliability, and safety of these drugs in both uncomplicated and severe malaria. The parent drug artemisinin was largely replaced by DHA and its derivatives artesunate and artemether, which have greater antimalarial activity (Fig. 2). Initially artemisinin and its derivatives were used as monotherapies, as it became gradually accepted that antimalarials, like antituberculosis and antiretroviral drugs, should be used in combination (5, 6). A semisynthetic artemisinin derivative (artelinate) was developed, but was not taken beyond animal studies, and another (artemisone) is under development. Artiflene, a structurally dissimilar peroxide derivative of another Chinese plant (yinghaosu) was developed in the early 1990s and proved an effective antimalarial in clinical trials, but it was eventually abandoned because of high production costs and lack of evident advantages over the artemisinin derivatives.

Pharmacological Properties

Artemisinins kill nearly all of the asexual stages of parasite development in the blood (7), and also affect the sexual stages of *P. falciparum* (gametocytes), which transmit the infection to others (8), but they do not affect pre-erythrocytic development or the latent stages of *P. vivax* and *P. ovale* in the liver (the hypnozoites). The mechanism of action of artemisinins remains uncertain. The integrity of the endoperoxide bridge is necessary (but not sufficient) for antimalarial activity. Iron-dependent alkylation (principally by Fe^{++}) is a likely mode of action (9), and the sarcoplasmic endoplasmic

Hours of development
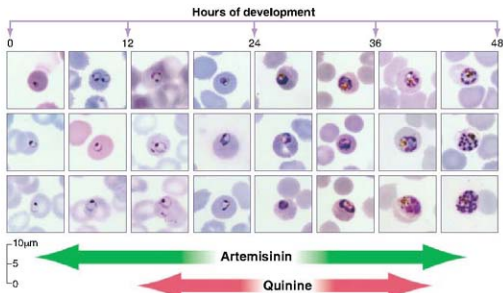
Fig. 1. The intraerythrocytic life cycle of *P. falciparum*. Parasitized red cells circulate for the first third of the 48-hour cycle and then sequester in capillaries and venules. Artemisinins inhibit development of a broader age range of the parasites than do quinine and other antimalarial drugs. The effect on the young rings prevents their development to the more pathological mature parasites that sequester.

Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, and Centre for Clinical Vaccinology and Tropical Medicine, Churchill Hospital, Oxford, UK. E-mail: nickw@tropmedres.ac

Fig. 2. Commercial cultivation of *Artemisia annua* (qinghao/huang hua hao) in China.

reticulum calcium adenosine triphosphatase (PfATPase 6) has been proposed as the primary target (*10*). The role of reactive decomposition intermediates such as carbon-centered free radicals remains controversial. The report that *P. falciparum* parasites from French Guiana with point mutations in the gene encoding PfATPase 6 were relatively resistant to artemether seemed to have fulfilled the molecular Koch's postulates for this target (*11*), but these findings have not yet been reproduced elsewhere. The synthetic peroxide trioxolanes, which are potent antimalarials, are 2 to 3 orders of magnitude less active in inhibiting PfATPase 6 (*12*).

Artemisinin's broad stage specificity of antimalarial action (Fig. 1) has two therapeutic consequences. Killing young circulating ring-stage parasites in *P. falciparum* infections results in a more rapid reduction in parasitaemia compared with other antimalarials (Fig. 4) and reduces considerably the number of parasites that mature to sequester in and block capillaries and venules (*13, 14*). This explains the rapidity of clinical responses and the life-saving benefit in severe malaria compared with quinine (which does not stop sequestration because it kills parasites only after they have matured and adhered to vascular endothelium). Reducing gametocyte carriage diminishes the transmission potential of the treated infection.

Artemisinin is eliminated by metabolic biotransformation, predominantly by CYP 2B6, to inactive metabolites. The artemisinins are weak inducers of some important drug-metabolizing enzymes and augment their own clearance (*15, 16*). After oral or parenteral administration, artemether, artemotil, and artesunate are all converted back rapidly to DHA in vivo, which is then eliminated by glucuronidation with an elimination half-life of ~1 hour, both in healthy volunteers and in patients with malaria (Fig. 5). The broad stage specificity of action ensures that a single daily

administration is sufficient for maximal killing of sensitive parasites. A 3-day artemisinin combination treatment (ACT) regimen provides antimalarial activity for two asexual parasite cycles and results in a reduction by a factor of about 100 million in parasite numbers within the infected patient—but this still leaves up to 100,000 parasites for the partner drug to remove, variably assisted by the immune response (*14*). The artemisinin component of the ACT therefore reduces the
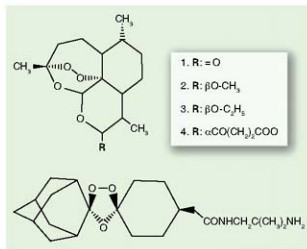
Fig. 3. Chemical structures. (**Top**) Artemisinin and its derivatives. 1, artemisinin; 2, artemether; 3, artemotil (arteether); 4, artesunate. (**Bottom**) The synthetic trioxolane OZ 277 (*51*).

probability that a mutant resistant to the partner drug would arise from the primary infection, and, if effective, the partner should kill any artemisinin-resistant parasite that arose.

Artemisinins also have clinically important activity against other parasites. They kill the younger stages of trematodes and are effective in the treatment of schistosomiasis and fascioliasis and in animal models of Clonorchis infections (*17, 18*). Their in vitro activity against other protozoa is

considerably less than against malaria parasites, although they might be of value in the treatment of African and South American trypanosomiasis (*19*). Artemisinins have anti-inflammatory properties and also inhibit angiogenesis and cell growth in several neoplastic cell lines, which suggests a potential role in cancer chemotherapy (*20, 21*).

Artemisinin Combination Treatment

When artemisinins are given alone, 7-day regimens are required to maximize cure rates. Adherence with 7-day treatment courses is poor, so the combination partner in ACTs is usually a slowly eliminated antimalarial drug. This allows a complete treatment course to be given in 3 days (*22*). The first ACT to be evaluated systematically was artesunate-mefloquine (*3, 23*). This was deployed in 1994 on the northwest border of Thailand—an area of mefloquine-resistant falciparum malaria—and has retained efficacy over the subsequent 14 years (*24*). The first fixed-dose ACT (artemether-lumefantrine) followed soon afterward. Other ACTs have combined artesunate with existing drugs (sulfadoxine-pyrimethamine or amodiaquine) (*6*). Their evaluation (*25*) coincided with increasing realization that, whereas mortality from most infectious diseases (with the exception of HIV-AIDS) was declining, malaria mortality was rising (*26, 27*). This was attributed directly to the continued use of increasingly ineffective antimalarial drugs, mainly chloroquine and sulfadoxine-pyrimethamine (SP). Where the partner drug had not already fallen to resistance, the new ACTs were effective and well tolerated, but they were more expensive than the failing monotherapies.

Malaria affects the poorest people in the poorest countries. Many affected people can barely afford the US $.10 equivalent to buy ineffective chloroquine. Malaria control activities are heavily dependent on aid, and initially this was not forthcoming for the ACTs, so the necessary policy changes were delayed (*28*). In the last few years, however, there has been a massive increase in donor funding, spearheaded by the Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM). Politicians in temperate countries, whose constituents have long been free from the burden of malaria, began to appreciate the humanitarian and economic burden that malaria imposes on the developing world. At the end of 2007, Bill and Melinda Gates called for a sustained initiative to eliminate malaria. It has taken a long time for the world to think positively again about conquering malaria (*24, 29, 30*) after the painful failure of the first global eradication campaign to eliminate the disease from the tropics. Whether malaria can and will be eliminated from high-transmission areas, where the vectorial capacity is enormous, remains to be seen, but most

agree that falciparum malaria elimination in many low and unstable transmission settings is feasible (these include much of Asia and South America and large areas of Africa). Providing highly effective treatment is an essential component of malaria control and is required for elimination.

ACTs are now recommended by the WHO as the first-line treatment for all falciparum malaria in malaria endemic countries. WHO recommends aiming for cure rates of 95%, assessed at 28 days, and changing antimalarial treatment policy if cure rates are less than 90% (31). Artesunate-mefloquine and artemether-lumefantrine are reliably efficacious everywhere except Western Cambodia (32). Artesunate-amodiaquine and artesunate-SP give cure rates over 90% when cure rates with the corresponding monotherapies exceed 80%. In low-transmission settings, where symptomatic infections constitute a major source of transmission, ACTs reduce gametocyte carriage and, if widely deployed, reduce the incidence of falciparum malaria (24, 29, 30). The ACTs are also highly effective against infections caused by *P. vivax*, *P. malariae*, and *P. ovale* (except for artesunate-SP against antifol-resistant *P. vivax*), but most people who need ACTs still do not receive them.

Severe Malaria

In the 1990s, several randomized trials were conducted that compared parenteral artemether with quinine. An individual patient data meta-analysis of these trials did not show an overall difference in mortality between the two treatments, although in the prospectively defined subgroup of Southeast Asian adults, mortality was significantly lower in artemether recipients (33). The oil-based artemether is absorbed slowly and erratically from intramuscular injection sites (Fig. 5), particularly in severely ill patients (34, 35). This pharmacokinetic disadvantage may have offset the intrinsic pharmacodynamic advantage of the artemether in killing malaria parasites. By contrast, the water-soluble artesunate can be given by bolus intravenous injection and is absorbed rapidly and reliably after intramuscular injection. In the largest prospective randomized study in severe malaria (which enrolled 1461 patients), conducted in Southeast Asia, artesunate reduced the mortality of severe malaria from 22 to 15%—a 35% reduction (36) (Fig. 6). Artesunate is now the treatment of choice for severe falciparum malaria in areas of low transmission (31). However, most deaths from severe malaria occur in or near home. If treatment were available close to home, then the lethal progression to severe malaria might be halted. A formulation of artesunate for rectal administration with ~50% bioavailability in acute malaria has been developed (37) that can be administered easily by a village health worker. In very large field studies, it has proved safe, well tolerated, and effective.

Safety and Efficacy

The artemisinins have proved to be well tolerated and safe drugs. They temporarily suppress erythropoiesis, but this does not cause significant anemia. The only potentially serious adverse effects are relatively unusual hypersensitivity reactions (~1 in 3000 treatments) that manifest initially as urticaria. In experimental animals, sustained central nervous system exposure to high concentrations of the artemisinins produces an unusual selective pattern of damage to certain brain-stem nuclei, particularly those involved with hearing and balance (38). Experimental neurotoxicity is therefore greater after intramuscular administration of high-dose oil-based artemether and artemotil, which are released slowly from the injection site, than with the same doses given orally (despite much higher peak concentrations) or the water-soluble artesunate given by any route (39). With the exception of a much-debated case series (40), extensive clinical and neurophysiological studies and a small series of neuropathological evaluations have provi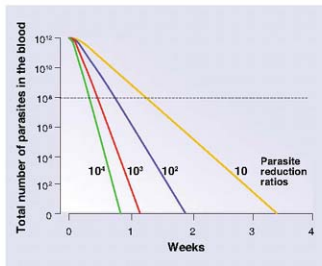ded no evidence of neurotoxicity in humans (41–43). Although concerns over potential neurotoxicity have receded, the animal data do argue against using artemisinins for prophylaxis or giving long courses of high-dose intramuscular artemether. The one unresolved safety concern is in the first trimester of pregnancy. When experimental animals are exposed to artemisinins during early pregnancy, fetal resorption may result. This results from temporary suppression of fetal erythropoiesis caused by depletion of fetal erythroblasts (44, 45). Whether artemisinins could be teratogenic in humans has not been established. So, although there is increasing evidence of safety for the artemisinins in the second and third trimesters, these drugs are not recommended in the first trimester except in severe malaria, where their life-saving benefit to the mother outweighs the putative risks to the fetus (31).

In uncomplicated malaria, the safety, tolerability, and efficacy of ACTs is determined largely by the partner drug, because the artemisinin component is very well tolerated and provides a fixed antimalarial effect. The WHO currently recommends four ACTs (artesunate-sulfadoxine-pyrimethamine, artesunate-amodiaquine, artesunate-mefloquine and artemether-lumefantrine) and, to ensure adherence to combination treatment, encourages the use of fixed-dose combinations (FDCs) (31). Dihydroartemisinin-piperaquine is another highly effective FDC that is already used in several countries. Artesunate-pyronaridine is an FDC ACT in the later stages of development. Whereas ACTs containing mefloquine, lumefantrine, and piperaquine are effective nearly everywhere in the world, high levels of resistance to pyrimethamine, amodiaquine, and chlorproguanil, particularly in Asia and South America, limit the usefulness of ACTs containing these drugs.

Resistance

Because current antimalarial treatment is so dependent on artemisinins, there has been considerable concern that resistance would emerge to this class of drugs as it has to other classes of antimalarials. Reassuringly, it has proved difficult to induce stable high-level resistance to artemisinins in the laboratory. Rodent malarias with reductions in susceptibility by factors of 5 to 10 have been selected, and similar reductions in susceptibility have been obtained by selection in *P. falciparum* (46). Amplification of *Pfmdr1* is associated with relatively small but significant reductions in susceptibility to artemisinins. Most isolates of *P. falciparum* are extremely sensitive to artemisinin and its derivatives; in vitro median inhibitory concentration (IC$_{50}$) values for dihydroartemisinin and artesunate are typically below 10 nmol/L, with values for artemether that are usually 2 to 3 times as high, and for artemisinin 5 to 10 times as high. Jambou *et al.* reported isolates from French Guiana with artemether IC$_{50}$ values up to 117 nmol/L associated with mutations in PfATPase 6. Mutations in this gene are not associated with resistance in field isolates from elsewhere or in the laboratory lines selected for resistance (11). Yang *et al.* reported reduction by a factor of 3.3 in the susceptibility of *P. falciparum* to artesunate between 1988 and

Fig. 4. The pharmacodynamic properties of the antimalarial drugs in vivo assessed in terms of parasite clearance (14). Shown are infections of 10^{12} parasites corresponding to ~2% parasitemia in an adult. The horizontal dashed line is the threshold for microscopic detection. Weak antimalarials, such as many antibiotics with antimalarial properties, produce parasite reduction ratios (PRR) of about a factor of 10 per asexual cycle, and therefore take weeks to eradicate an infection. Most antimalarials have PRR values of 100 to 1000 per cycle. The artemisinin derivatives are the most potent antimalarial drugs, with PRR values of ~10,000 per cycle, and therefore take only three to four cycles (6 to 8 days) to remove all the parasites from the blood.
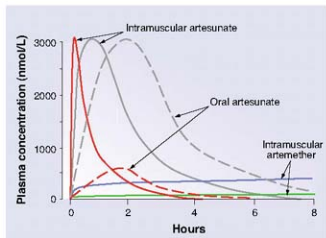
Fig. 5. Comparative pharmacokinetics of intramuscular (i.m.) artesunate (2.4 mg/kg), oral artesunate (4 mg/kg), and i.m. artemether (3.2 mg/kg) in acute malaria. The plasma concentrations of the parent compound are shown as a red line for i.m. artesunate, a red dashed line for oral artesunate, and a blue line for i.m. artemether. The common active metabolite dihydroartemisinin is shown as gray or green lines.

1999 in Yunnan, Southwest China (47). This is a region where artemisinins have been used extensively for more than 20 years. More recently, studies from the Thai-Cambodian border have demonstrated reduced susceptibility to artemisinins both in vitro and in vivo (48). This is an area notorious as the epicenter from which chloroquine-resistant and later multidrug-resistant *P. falciparum* spread 50 years ago. Treatment failure rates after artesunate-mefloquine and artemether-lumefantrine in western Cambodia often exceed 10% (32), consistently higher than anywhere else in the world. Although *P. falciparum* is relatively resistant to mefloquine and lumefantrine in this area, and this certainly contributes to the high failure rates, there is concern that sensitivity to artemisinins may have declined as well. Artemisinins have been used in a variety of formulations and doses in western Cambodia for approximately 30 years, providing a cumulative exposure that may be more than anywhere else in the world. Although in vitro susceptibility tests indicate a relative reduction in artemisinin susceptibility, compared with most other regions of the world, the parasite isolates studied from this region are not highly resistant (i.e., reductions in susceptibility by no more than a factor of 10). They are not more resistant in vitro than parasites isolated on the eastern and western borders of Thailand, but in vivo the parasite clearance times in western Cambodia are consistently longer than elsewhere (49). Rapid parasite clearance is the pharmacodynamic hallmark of the artemisinins, reflecting their unique ac-

tivity against young ring-stage parasites (14). Thus, there is a discrepancy between the in vitro and in vivo findings. Intensive studies are under way to assess the geographic extent of the problem, to characterize the in vivo and in vitro responses further and, if possible, to identify the molecular basis of the artemisinin-tolerant phenotype (49).

In assessing possible artemisinin resistance, specific features of the therapeutic response to these drugs needs to be considered. First, if the artemisinins are given alone for 7 days to nonimmune patients, ~10% of patients fail treatment, yet when the recrudescent parasites are compared with parasites isolated from successfully treated infections, they are not more resistant.

The persistence of temporarily growth-arrested intraerythrocytic merozoites or young trophozoites, which can "awaken" from dormancy days or weeks later, provides a plausible explanation for this result (50). The biological basis for this phenomenon is not understood. Second, apparently slow parasite clearance has been observed in a small proportion of patients in most trials conducted in nonimmune patients. In patients who have undergone a previous splenectomy, dead intraerythrocytic parasites may persist for up to a month in the circulation, so reduced splenic phagocytic function might explain slow

tests may be insensitive to the early development of artemisinin resistance.

Reducing Costs

Artemisinin is extracted from the leaves of *Artemisia annua*, which must be planted each year. China and Vietnam are major producers, and in recent years production in Africa has been increasing. Temperature, humidity, and soil type all affect yields considerably. Artemisinin is reduced in a potentially explosive reaction to dihydroartemisinin, which can then be converted to more heat-stable ethers (artemether, artemotil) or the succinate derivative (artesunate). While the market was small, prices were volatile, with prices for artemisinin ranging from US $350 to $1700 per kilogram on world markets. In recent years, there has been an increase in agricultural production and improvements in horticulture and yields. The number of companies extracting artemisinin has risen from less than 20 to more than 100. This has resulted in a reduction in the market price of artemisinin (52) and in retail costs, which can be less than US $1 for an adult treatment course. This is still approximately ten times as expensive as its synthetic predecessor chloroquine, and even chloroquine was unaffordable for many, so there have been considerable price barriers to widespread use. Current initiatives to reduce cost include (i) improvement in horticulture by fast-track selective breeding without genetic modification (53), (ii) microbial synthesis of the artemisinin precursor artemisinic acid followed by chemical synthesis of artemisinins (54), and (iii) development of fully synthetic antimalarial peroxides (55).

The microbial production of artemisinin is expected to begin in 2010 with goals of producing artemisinin at US $100 per kilogram and enough medicine for 200 million treatments each year. A large number of antimalarial peroxides have been synthesized over the past 15 years, but only one (OZ 277/ RBx-11160, Arterolane) entered clinical development (55). This trioxolane compound was orally bioavailable and highly potent, but its development was discontinued before phase 3 clinical trials because of its instability in blood. More stable peroxides are now in development, and hybrid trioxane-aminoquinoline molecules have been developed (trioxaquines), which combine two mechanisms of action in a single molecule (56). These are undoubtedly promising developments, and it is likely that one or more of these candidates will eventually realize the objective of an extremely low-cost, fully synthetic peroxide antimalarial.

Increasing Access

Access to health care through the public sector in many tropical countries is inadequate, and for people with little or no money, malaria medicines may be unaffordable. Most febrile illnesses are treated empirically without any diagnostic procedure. Antimalarials are widely used commodities that can usually be purchased from shops, in
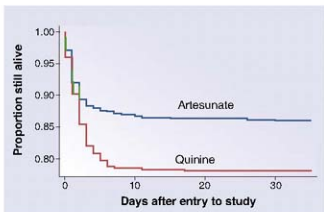
Fig. 6. Treatment with intravenous artesunate was associated with a 35% reduction in mortality from severe malaria compared with intravenous quinine in a randomized trial of 1461 adults and children in Southeast Asia (36). The vertical axis shows the proportion of patients surviving.

parasite clearance in some cases (51). Third, current in vitro susceptibility tests assess maturation to mature schizonts, or the production of proteins or nucleic acid, which is maximal in mature trophozoites, whereas accelerated parasite clearance reflects drug effects on young circulating ring stages of the parasite. Current in vitro

the market, or from itinerant drug sellers, although in remote communities access to drugs is usually poor. The doses recommended are commonly incorrect, the quality of the drugs is often poor, and counterfeits are a major problem (57). Ineffective drugs are often purchased because they are less expensive, or an incomplete course of treatment is taken so that some tablets can be retained for the next febrile illness. The net result is both over- and undertreatment and increased selection pressure for the emergence and spread of resistance. Strengthening health service infrastructures and increasing the provision of diagnostics and effective treatment are vital components of current control efforts, but to reach most people in need, the availability of effective, affordable ACTs through the private sector also needs to be strengthened. In many malaria-endemic countries, this is the main source of treatment. The Institute of Medicine reviewed the economics of antimalarial drugs and in 2004 concluded that the most effective way of increasing access to antimalarial drugs would be to ensure that quality-assured effective drugs (currently ACTs) were provided free to the public sector and were also made available in the market place at a price close to that of the now largely ineffective chloroquine (about US $.10 to $.20 for a treatment) (58). This would have the added advantages of reducing the financial incentives for incomplete treatment, outcompeting ineffective drugs while driving out more costly and less effective artemisinin monotherapies, and reducing considerably the incentives to increasingly sophisticated and pervasive counterfeiters (59). It was proposed that a central subsidy be provided by international donors. This concept was taken forward by the Roll Back Malaria (RBM) Finance and Resource Working Group and the World Bank. In November 2007, the RBM board endorsed the further development of an Affordable Medicines Facility for malaria (AMFm), a groundbreaking initiative to improve access to safe, effective, and affordable malaria medicines (60). It is expected that the AMFm will require US $1.4 to 1.9 billion in funding over 5 years. This can be compared with estimates of the costs of malaria: US $12 billion a year in lost productivity in Africa alone and, in countries with a very heavy malaria burden, as much as 40% of public health expenditure, 50% of inpatient admissions, and up to 60% of outpatient visits (61). It is definitely a price worth paying.

References and Notes

1. Qinghaosu Antimalarial Coordinating Research Group, *Chin. Med. J. (Engl.)* **92**, 811 (1979).
2. G. Schmid, W. Hofheinz, *J. Am. Chem. Soc.* **105**, 624 (1983).
3. J. B. Jiang et al., *Lancet* **2**, 285 (1982).
4. G. Q. Li et al., *J. Tradit. Chin. Med.* **2**, 125 (1982).
5. W. Peters, *Pharmacol. Ther.* **47**, 499 (1990).
6. International Artemisinin Study Group, *Lancet* **363**, 9 (2004).
7. F. ter Kuile et al., *Exp. Parasitol.* **76**, 86 (1993).
8. P. Q. Chen et al., *Chin. Med. J. (Engl.)* **107**, 709 (1994).
9. R. K. Haynes et al., *ChemMedChem* **2**, 1480 (2007).
10. U. Eckstein-Ludwig et al., *Nature* **424**, 957 (2003).
11. R. Jambou et al., *Lancet* **366**, 1960 (2005).
12. A. C. Uhlemann et al., *Antimicrob. Agents Chemother.* **51**, 667 (2007).
13. R. Udomsangpetch et al., *J. Infect. Dis.* **173**, 691 (1996).
14. N. J. White, *Antimicrob. Agents Chemother.* **41**, 1413 (1997).
15. M. Ashton et al., *Clin. Pharmacol. Ther.* **63**, 482 (1998).
16. S. Asimus et al., *Fundam. Clin. Pharmacol.* **21**, 307 (2007).
17. J. Utzinger et al., *Lancet* **355**, 1320 (2000).
18. J. Karbwang et al., *Antimicrob. Agents Chemother.* **51**, 1096 (2007).
19. Y. V. Mishina, S. Krishna, R. K. Haynes, J. C. Meade, *Antimicrob. Agents Chemother.* **51**, 1852 (2007).
20. H. Xu et al., *Rheumatology (Oxford)* **46**, 920 (2007).
21. I. Nakase, H. Lai, N. P. Singh, T. Sasaki, *Int. J. Pharm.* **354**, 28 (2007).
22. M. Al. Nyunt, C. V. Plowe, *Clin. Pharmacol. Ther.* **82**, 601 (2007).
23. F. Nosten et al., *J. Infect. Dis.* **170**, 971 (1994).
24. V. I. Carrara et al., *PLoS Med.* **3**, e183 (2006).
25. International Artemisinin Study Group, *Lancet* **363**, 9 (2004).
26. E. L. Korenromp et al., *Lancet Infect. Dis.* **3**, 349 (2003).
27. A. D. Lopez et al., *Lancet* **367**, 1747 (2006).
28. N. J. White et al., *Lancet* **353**, 1965 (1999).
29. K. I. Barnes et al., *PLoS Med.* **2**, e330 (2005).
30. A. Bhattarai et al., *PLoS Med.* **4**, e309 (2007).
31. World Health Organization, *The Treatment of Malaria* (WHO, Geneva, 2006).
32. M. B. Denis et al., *Trop. Med. Int. Health* **11**, 1360 (2006).
33. Artemether-Quinine Meta-analysis Study Group, *Trans. R. Soc. Trop. Med. Hyg.* **95**, 637 (2001).
34. S. A. Murphy, *Trans. R. Soc. Trop. Med. Hyg.* **91**, 331 (1997).
35. T. T. Hien, *Antimicrob. Agents Chemother.* **48**, 4234 (2004).
36. A. Dondorp et al., *Lancet* **366**, 717 (2005).
37. K. I. Barnes, *Lancet* **363**, 1598 (2004).
38. T. G. Brewer et al., *Trans. R. Soc. Trop. Med. Hyg.* **88** (Suppl. 1), S33 (1994).
39. R. F. Genovese, D. B. Newman. *Arch. Toxicol.*; published online 31 October 2007; doi: 10.1007/s00204-007-0252-z
40. S. Toovey, A. Jamieson, *Trans. R. Soc. Trop. Med. Hyg.* **98**, 261 (2004).
41. E. Kissinger et al., *Am. J. Trop. Med. Hyg.* **63**, 48 (2000).
42. M. van Vugt, *Am. J. Trop. Med. Hyg.* **62**, 65 (2000).
43. T. T. Hien et al., *Lancet* **362**, 295 (2003).
44. M. Longo et al., *Reprod. Toxicol.* **21**, 83 (2006).
45. T. E. White et al., *Birth Defects Res. B Dev. Reprod. Toxicol* **77**, 413 (2006).
46. P. Hunt et al., *Mol. Microbiol.* **65**, 27 (2007).
47. H. Yang et al., *Trans. R. Soc. Trop. Med. Hyg.* **97**, 226 (2003).
48. *Wkly. Epidemiol. Rec.* **82**, 360 (2007).
49. WHO Global Malaria Programme, informal consultation on global malaria control and elimination: Containment of artemisinin resistance at the Thai-Cambodia border, Geneva, in preparation.
50. D. E. Kyle, presentation at the Keystone Symposia Global Health Series, 28 to 31 January 2007, Tahoe City, California.
51. K. Chotivanich, *J. Infect. Dis.* **185**, 1538 (2002).
52. J. M. Kindermans, J. Pilloy, P. Olliaro, M. Gomes, *Malaria J.* **6**, 125 (2007).
53. D. K. Ro et al., *Nature* **440**, 940 (2006).
54. M. C. Chang et al., *Nat. Chem. Biol.* **3**, 274 (2007).
55. J. L. Vennerstrom et al., *Nature* **430**, 900 (2004).
56. F. Benoit-Vical et al., *Antimicrob. Agents Chemother.* **51**, 1463 (2007).
57. P. N. Newton et al., *PLoS Med.* **3**, e197 (2006).
58. Institute of Medicine, *Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance* (National Academies Press, Washington, DC, 2004).
59. P. N. Newton et al., *PLoS Med.* **5**, e32 (2008).
60. http://go.worldbank.org/E7PZNUSD5O, 29 November 2007.
61. Roll Back Malaria Partnership, http://rbm.who.int, accessed 1 January 2008.
62. I am very grateful to my colleagues in Thailand and Vietnam whose work is presented here, and to P. Olliaro, R. Price, and D. Kyle for advice. I am a Wellcome Trust Principal Fellow, and I am Chairman of the WHO Antimalarial Treatment Guidelines Committee.

The Movement of Aquatic Mercury Through Terrestrial Food Webs

Daniel A. Cristol,* Rebecka L. Brasso, Anne M. Condon, Rachel E. Fovargue, Scott L. Friedman, Kelly K. Hallinger, Adrian P. Monroe, Ariel E. White

Mercury is a persistent contaminant that biomagnifies up the food web, causing mortality, reproductive failure, and other health effects in predatory wildlife and humans (1, 2). From 1930 to 1950, industrial mercuric sulfate entered the South River, a tributary of the Shenandoah River in Virginia (United States) (3). To determine whether this mercury contamination had moved into the adjacent terrestrial food web, we analyzed total mercury concentrations in blood from adults of 13 terrestrial-feeding bird species breeding within 50 m of the river (4). We compared these to breeding adults of the same species at uncontaminated reference sites (Fig. 1A). Twelve individuals had significantly higher mercury than that found in individuals from reference sites (Fig. 1B and table S1). We simultaneously sampled adults of five bird species with direct dietary connections to the aquatic mercury; their mercury concentrations were also significantly elevated relative to reference birds (Fig. 1B and table S1). With the exception of a duck species, all aquatic birds had blood mercury ranging from 2 to 4 parts per million (ppm) (wet weight), about the level of adverse effects (5). Two terrestrial songbird species, a wren and a vireo, had the highest blood mercury levels of any species, and a majority of terrestrial-feeding birds were in the same range as aquatic-feeding species.

We simultaneously sampled feathers of three nonmigratory terrestrial species that had molted locally about a year earlier. Whereas blood mercury reflects recent dietary intake, feathers indicate cumulative exposure; mercury can only enter feathers while they are growing and connected to the blood supply. Wrens (*Thyrothorus ludovicianus*) averaged 8.76 ± 6.46 (SD) ppm (n = 6), compared with 5.63 ± 2.12 ppm for owls (n = 5) and 2.38 ± 1.31 ppm for woodpeckers (n = 10), all about twice the level of blood mercury.

We collected prey delivered to nestlings of three songbird species at all of the contaminated study sites. None of the diet items was aquatic in origin. The most prevalent items in all three bird species' diets were members of the orders Araneae (spiders), Lepidoptera (moths or caterpillars), and Orthopterae (grasshoppers). Together these comprised >80% of the biomass (fresh weight) delivered to nestlings (Fig. 1C). All three terrestrial birds ate diets of ~20 to 30% spider biomass. Average total mercury concentration in spiders (n = 101) was 1.24 ± 1.47 ppm dry weight, compared with only 0.38 ± 2.08 ppm for lepidopterans (n = 137) and 0.31 ± 1.22 ppm for orthopterans (n = 50). Invertebrates collected similarly on reference sites contained negligible mercury (spiders, 0.05 ± 0.03 ppm, n = 25; lepidopterans, 0.02 ± 0.64 ppm, n = 20; orthopterans, all below detection limit, n = 6).

Much of the mercury in the spiders was in the form of highly bioavailable methylmercury (49 ± 21%, n = 30), whereas invertebrates lower in the food web had less methylmercury (lepidopterans 24 ± 20%, n = 22; orthopterans 38 ± 24%, n = 35), consistent with other reports (6). The spiders at our study site had higher total mercury than fish collected from kingfishers at the site (contaminated fish 0.73 ± 1.76 ppm dry weight, n = 21; reference

fish, 0.15 ± 0.09 ppm, n = 18), consistent with our observation that some terrestrial songbirds had higher mercury levels than fish-eating kingfishers.

These results show that any risk from mercury faced by the river's aquatic birds exists for birds in adjacent terrestrial habitats as well. By preying on predatory invertebrates such as spiders, songbirds increased the length of their food chains, increasing the opportunity for biomagnification. To assess the potential risk of aquatic mercury moving into terrestrial food webs, it is imperative to determine whether the methylmercury we observed in predatory invertebrates was transported directly to the terrestrial food web by emergent aquatic insects or had been deposited on the floodplain during historical floods.

References and Notes

1. A. M. Scheuhammer, M. W. Meyer, M. B. Sandheinrich, M. W. Murray, *Ambio* 36, 12 (2007).
2. D. Mergler et al., *Ambio* 36, 3 (2007).
3. L. J. Carter, *Science* 198, 1015 (1977).
4. Materials and methods are available on Science Online.
5. D. Evers et al., *Ecotoxicology* 17, 69 (2008).
6. T. A. Haines, T. W. May, R. T. Finlayson, S. E. Mierzykowski, *Environ. Monit. Assess.* 86, 211 (2003).
7. Funded by E. I. DuPont de Nemours and Company, Vice Provost for Research of College of William and Mary, and NSF UBM 0436318. J. Schmerfeld, U.S. Fish and Wildlife Service, suggested the study. R. Taylor analyzed avian mercury, and C. Shade analyzed methylmercury. We thank A. Adkhipov, D. Hawley, S. Hoskin, M. Howie, R. Jefferson-George, S. Koebley, M. Leandre, L. Savoy, J. Spahr, and C. Ziegenfus for fieldwork and D. Evers, T. Meier, M. Newman, and the entire South River Science Team for logistical support.
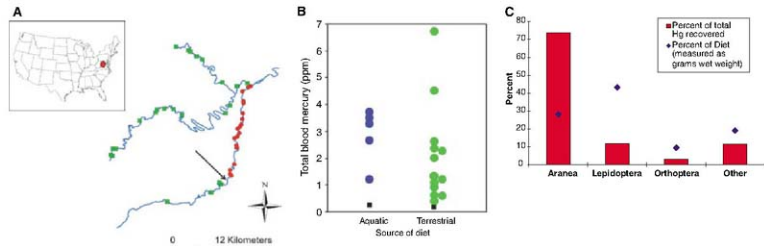
Fig. 1. (**A**) Map of study sites in which red circles indicate contaminated collection sites and green squares indicate uncontaminated sites. Arrow indicates mercury source. (**B**) Mean blood mercury for bird species primarily feeding on aquatic (blue) or terrestrial (green) organisms. Black squares are average mercury levels for same species at reference sites. (**C**) Percentage of diet composed of each prey type (blue diamonds) and mercury intake (red bars) for two wren species.

Phytoplankton Calcification in a High-CO₂ World

M. Debora Iglesias-Rodriguez,[1]* Paul R. Halloran,[2]* Rosalind E. M. Rickaby,[2]
Ian R. Hall,[3] Elena Colmenero-Hidalgo,[3]† John R. Gittins,[1] Darryl R. H. Green,[1]
Toby Tyrrell,[1] Samantha J. Gibbs,[1] Peter von Dassow,[4] Eric Rehm,[5]
E. Virginia Armbrust,[5] Karin P. Boessenkool[3]

Ocean acidification in response to rising atmospheric CO₂ partial pressures is widely expected to reduce calcification by marine organisms. From the mid-Mesozoic, coccolithophores have been major calcium carbonate producers in the world's oceans, today accounting for about a third of the total marine CaCO₃ production. Here, we present laboratory evidence that calcification and net primary production in the coccolithophore species *Emiliania huxleyi* are significantly increased by high CO₂ partial pressures. Field evidence from the deep ocean is consistent with these laboratory conclusions, indicating that over the past 220 years there has been a 40% increase in average coccolith mass. Our findings show that coccolithophores are already responding and will probably continue to respond to rising atmospheric CO₂ partial pressures, which has important implications for biogeochemical modeling of future oceans and climate.

The climatological and ecological impacts of elevated atmospheric CO₂ partial pressures (P_{CO_2}) are two of the most pressing environmental concerns of the present. One consequence of increasing P_{CO_2} in seawater is the formation of carbonic acid (H₂CO₃), which causes acidification. Carbonic acid combines with carbonate ions (CO₃²⁻) and water molecules to form bicarbonate ions (HCO₃⁻), reducing [CO₃²⁻] and the ocean's saturation state with respect to calcite (Ω-cal), the form of calcium carbonate (CaCO₃) produced by coccolithophores. Elevated P_{CO_2} also causes an increase in [HCO₃⁻], the source of carbon for calcification in coccolithophores (Ca²⁺ + 2HCO₃⁻ → CaCO₃ + CO₂ + H₂O) (1). Thus, calcification is probably affected by increasing P_{CO_2}. The precipitation from seawater of CaCO₃, a basic substance, lowers pH. For this reason, and because a greater fraction of dissolved inorganic carbon {DIC, the sum of HCO₃⁻, CO₃²⁻, and aqueous [CO₂(aq)]} is present as CO₂(aq) at low pH, the formation of CaCO₃ in seawater stimulates an increase in the concentration of CO₂(aq) and promotes its outgassing. Consequently, a decrease in marine calcification without a concomitant decrease in organic carbon export would lead to an increased drawdown of atmospheric CO₂.

Recent evidence suggests that the increased absorption of CO₂ by the oceans, as a result of anthropogenic CO₂ release, will result in decreased calcification by corals (2), foraminifera (3), and coccolithophores (4–6). However, it has recently been shown that different coccolithophore species exhibit different calcification responses. Under increased P_{CO_2}, a decrease in calcification has been observed for *Emiliania huxleyi* and *Gephyrocapsa oceanica* (4–6); a negligible calcification change with rising P_{CO_2} for *Coccolithus pelagicus* (7); and an increase followed by a decrease in calcification with rising P_{CO_2}, with respect to present-day P_{CO_2}, for *Calcidiscus leptoporus* (7). Most of these experiments used semicontinuous cultures, in which the carbonate system was modified by the addition of acid and/or base to control pH (4, 5, 7). Seawater pH controls the relative proportion of the carbonate species while the concentration of DIC remains constant. A more realistic representation of the ocean response to anthropogenic change is the bubbling of CO₂-enriched air through the seawater, both elevating [DIC] and decreasing pH. Recent studies with various organisms show calcification to be largely controlled by Ω-cal, rather than pH alone (7, 8), and Ω-cal is controlled by both [DIC] and pH. Between the years 1800 and 2100, seawater pH is likely to fall from 8.2 to 7.8 (9). Achieving the required pH by CO₂ bubbling induces a greater percentage increase in [HCO₃⁻] than when the same pH reduction is achieved through acid addition (which does not affect [DIC]). Therefore, to investigate calcification under future CO₂ scenarios, it is important to correctly simulate [HCO₃⁻].

We designed experiments that accurately represent projections of the future carbonate system, and assessed the natural response of coccolithophores in the sedimentary record to infer these relationships over the past two centuries. Labora-

tory experiments tested the effect of increasing P_{CO_2} on calcification and other physiological parameters in the globally important coccolithophore species *E. huxleyi*. We then considered the laboratory results in the context of a field study, using sediment material from the box core RAPID 21-12-B (10) to examine assemblagewide changes in coccolith mass over the past ~220 years in response to anthropogenic CO₂ release.

Culture experiments. We conducted batch incubations with exponentially growing cells of the coccolithophore species *E. huxleyi* (11). Commercially manufactured air containing different P_{CO_2} was bubbled through the culture medium to adjust the P_{CO_2} of cultures from preindustrial levels [280 parts per million by volume (ppmv) of CO₂] up to the level predicted by one scenario for the end of the 21st century (750 ppmv of CO₂) (12). Our results suggest a doubling of particulate inorganic carbon (PIC) and particulate organic carbon (POC) production at 750 ppmv of CO₂. Between 280 and 490 ppmv, carbon metabolism remained broadly similar. In contrast, between 490 and 750 ppmv, both cellular PIC and POC and their production rates increased significantly (Fig. 1 and table S1). Growth rates were substantially lower at 750 ppmv of CO₂ as compared with 280, 300, and 490 ppmv of CO₂ (Fig. 1 and table S1). In parallel to the increases in POC and PIC production, analyses of particle counts and volumes (Coulter counter and flow cytometry analysis) were conducted in a subset of experiments. These analyses demonstrated that the volumes of both coccospheres (protoplast and calcium carbonate plates or coccoliths) and coccoliths increased with rising P_{CO_2}, following a similar trend in PIC and POC (Fig. 2 and Table 1). The range of coccolith volumes is comparable to that reported in response to changing nutrient availability and salinity (13). Flow cytometry data indicated that the PIC increase in the medium under high P_{CO_2} was due to both an increase in the volume of calcite within the coccospheres and an increase in the production of detached coccoliths (Table 1). Scanning electron micrographs of cells did not reveal apparent malformation or dissolution of coccoliths under any of the experimental P_{CO_2} conditions (Fig. 2). Physiological changes related to increased PIC and POC production were not accompanied by alterations in the photochemical efficiency of photosystem II [the ratio of the variable-to-maximum fluorescence (Fv:Fm) ~ 0.48] (14), assessed using fast repetition rate fluorometry (FRRF) (14), indicating that cells remained "photosynthetically healthy" in all experiments (Fig. 1).

A key factor determining whether coccolithophore production represents a net source or sink of CO₂ to the atmosphere is whether the calcification-to-photosynthesis ratio is greater or less than 1.5 (15, 16). The coincident increase in both PIC and POC production per cell in all the P_{CO_2} treatments resulted in a stable PIC:POC ratio of less than 1, although interactions with other climate-driven parameters may affect the observed trends. Our

[1]National Oceanography Centre, Southampton, University of Southampton Waterfront Campus, European Way, Southampton SO14 3ZH, UK. [2]Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK. [3]School of Earth, Ocean and Planetary Sciences, Cardiff University, Main Building, Park Place, Cardiff CF10 3YE, UK. [4]Station Biologique de Roscoff, Place George Teissier, BP 74, 29682 Roscoff Cedex, France. [5]School of Oceanography, Box 357940, University of Washington, Seattle, WA 98195, USA.

*These authors contributed equally to this work.
†Present address: Departamento de Geología, Facultad de Ciencias, Universidad de Salamanca, 37008 Salamanca, Spain.

Fig. 1. Cellular PIC (**A**), POC (**B**), PIC production rates (**C**), POC production rates (**D**), C:N ratios (**E**), PIC:POC ratios (**F**), growth rates (**G**), and Fv:Fm (**H**) for *E. huxleyi* cultures under different P_{CO_2}. Each color represents one independent experiment. Significant increases with rising P_{CO_2} were observed for PIC ($F_{4,16} = 24.14$, $P < 0.001$), POC ($F_{4,9} = 10.23$, $P = 0.002$), PIC production ($F_{4,16} = 5.94$, $P = 0.004$), POC production ($F_{4,9} = 4.52$, $P = 0.028$), and growth rate ($F_{4,16} = 3.92$, $P = 0.021$) (table S1). Differences between the treatments of 600 and 750 ppmv of CO_2 were significant for PIC ($P = 0.002$) but nonsignificant ($P > 0.05$) for all other parameters. Cellular PIC and POC were comparable at 280, 300, and 490 ppmv of CO_2. Above 490 ppmv of CO_2, cellular PIC and POC increased significantly, by 80 and 90% respectively at 600 ppmv of CO_2, and by a further 48 and 45% respectively at 750 ppmv of CO_2. Variation in PIC and POC production rates between 280 and 490 ppmv was not significant (table S1). Between 490 and 600 ppmv of CO_2, PIC and POC production rates increased by approximately 44 and 81%, respectively, and these were approximately 30 and 18% higher at 750 than at 600 ppmv of CO_2. Growth rates were significantly lower at 750 ppmv of CO_2 as compared with 280, 300, and 490 ppmv of CO_2. Differences in PIC:POC under the different P_{CO_2} treatments were nonsignificant ($F_{4,9} = 1.22$, $P = 0.368$) (table S1). The C:N values increased from 6.8 at 280 ppmv of CO_2 to 8.3 at 750 ppmv of CO_2. Fv:Fm values were comparable in all P_{CO_2} treatments. The shaded area represents putative P_{CO_2} during the PETM [lower-end estimates of P_{CO_2} were based on stomatal index and boron isotopes, data compiled in (*38*)].
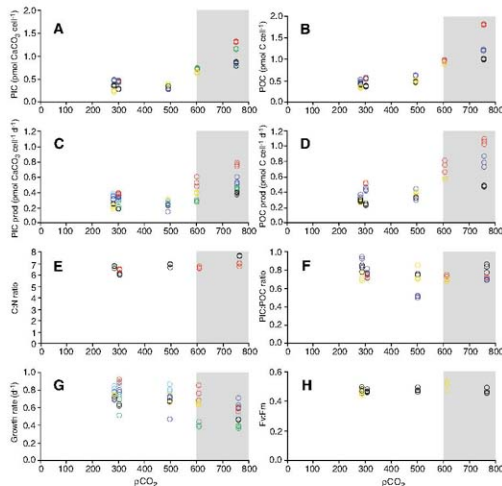
results suggest that levels of P_{CO_2} and Ω-cal corresponding to projections for the end of this century are unlikely to affect the metabolic balance between organic carbon fixation and calcite precipitation in *E. huxleyi*.

We measured the ratios of POC to particulate organic nitrogen (C:N) to assess whether the elemental composition of the organic material was additionally affected by changing P_{CO_2}. Variations in the elemental stoichiometry of phytoplankton are known to have an effect on trophic interactions, because the dietary value of prey items for marine zooplankton varies with the C:N ratio (*17*). Previous studies have reported changes in the elemental composition of diatoms in response to variations in P_{CO_2} (*18*). The C:N ratios in *E. huxleyi* increased from 6.8 to 8.3 with rising P_{CO_2} between 280 and 750 ppmv of CO_2 (Fig. 1). These results indicate that the P_{CO_2} could affect the grazing-selection pressure on phytoplankton, representing different "food" qualities. Grazing selection has many biogeochemical consequences and in particular implications for the export flux of carbon (*17*).

Our data show that Ω-cal ranged from 5.3 at 280 ppmv of CO_2 to 2.6 at 750 ppmv of CO_2, corresponding to an average total alkalinity of 2292 μeq liter^{-1} (Table 2). Ω-cal values were within the range of those for most of the upper-ocean regions, and well above 1, the threshold value below which dissolution would occur. In this pH range, less
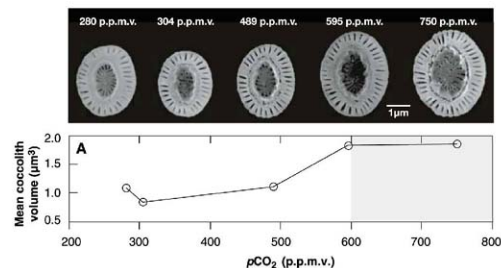
Fig. 2. Coccolith volume and $CaCO_3$ per cell. Increasing coccolith volume is closely coupled with increasing $CaCO_3$ per cell, indicating down-core measurement of coccolith mass to be representative of $CaCO_3$ production. Scanning electron microscope (SEM) images show typical coccoliths from each culture with P_{CO_2} values from 280 to 750 ppmv of CO_2, of where the measured volume was converted to length using the formula for a heavily calcified coccolith (*27*).

than 10% of the DIC in the medium was taken up by the proliferating cells (Table 2). Comparing these values with those in the corresponding blanks (without *E. huxleyi* cells) shows that cell physiology caused a shift in pH of less than 0.04 units in all experiments (Table 2). The pH values of the

cultures incubated at 280 and 750 ppmv of CO_2 ranged between 8.1 and 7.7 (corresponding to 9.5 μM CO_2 and 25.1 μM CO_2, respectively). These changes did not affect the photosynthetic health of cells (Fig. 1), which implies our pH conditions were within the tolerance levels of *E. huxleyi*. A

similar conclusion was reached in (19), where pH values within the range of those measured here did not suppress calcification. Our results are unlikely to be due to the physiological traits of a particular strain of *E. huxleyi*, because we observed the same effects on calcification and organic carbon production in another calcifying strain of *E. huxleyi* (61/12/4, Marine Biological Association, Plymouth, UK).

Down-core observations. In light of our laboratory results, which show a correlated increase in PIC and coccolith size with elevated P_{CO_2}, we investigated the response of a natural coccolithophore assemblage at high latitude to anthropogenic ocean acidification since the Industrial Revolution. We developed a method that can estimate the average mass of calcite per coccolith across multiple coccolithophore taxa (11). This technique was applied to material from the box core RAPID 21-12-B (57°27.09′N, 27°54.53′W), situated at 2630 m water depth in the subpolar North Atlantic. Core RAPID 21-12-B contains unprecedented open-ocean sedimentation rates of 2.3 mm year^{-1} spanning the time interval from 1780 to 2004 C.E. (10), which allows a detailed view of coccolith formation over the Anthropocene period, the period of anthropogenic CO_2 release.

Sediment was filtered at 10 μm to obtain the coccolith fraction, excluding larger carbonate grains (11) (fig. S1). The mass of calcite in two subsamples at each depth was measured in triplicate, and the number of $CaCO_3$ particles between 0.63 and 10 μm [reasonably assumed to be coccoliths (20, 21)] was counted nine times with an electrical resistance pulse detector (Coulter counter). Measurements were made before and after the addition of acid to account for the non-$CaCO_3$ component of the sediment. An upper detection limit of 10 μm was chosen to focus observations on particles with cohesive behavior and to avoid sampling the drift component of the sediment (22). This method excludes coccoliths with a diameter >10 μm. Only coccoliths of *C. pelagicus braarudii* were consistently >10 μm and were correspondingly excluded from the species counts. This approach measures the average mass of calcite per coccolith, which integrates any change in $CaCO_3$ mass due to variations in the assemblage and to intraspecies shifts in coccolith mass. To examine whether changes in species composition could account for the observed trend, coccolith assemblages were counted under a light microscope, following standard techniques for preparation by settling (23). No significant trend in species composition (Fig. 3) nor estimated species mass contribution (fig. S2) was observed. Dividing automated particle counts by sample weights before and after the removal of $CaCO_3$ by dissolution, and subtracting postdissolution measurements from predissolution measurements, not only rapidly provided average mass data on a large number of coccoliths (average sample counts were ~80,000 $CaCO_3$ particles), but was also sensitive to volume changes in the coccolith in any dimension.

The average mass of $CaCO_3$ per coccolith increased from 1.08×10^{-11} to 1.55×10^{-11} g between 1780 and the modern day (Fig. 4), with an accelerated increase over recent decades (fig. S3). Evidence is building that coccoliths are more resistant to dissolution than are planktonic foraminifera (24) and that they remain pristine when exposed to fluids in the pH range of 6 to 8 (25). In agreement with these observations, the absence of any down-core trend in coccolith species abundance in RAPID 21-12-B, despite the presence of taxa exhibiting a range of suscep-

Table 1. Coccosphere and coccolith volumes of *E. huxleyi* cells under different P_{CO_2} measured using a Coulter counter and flow cytometer. *t* tests of pairwise comparisons of the mean coccosphere and coccolith volumes measured by Coulter counter gave *P* values below 0.01 for all the pairwise comparisons. Side scatter (here in relative units, normalized to the side scatter of 3-μm internal bead standards) correlates strongly with the cellular calcification of *E. huxleyi* (39), whereas forward scatter correlates with coccosphere size. Comparison of forward-scatter volume before and after acidification indicates that the differences in volume among the different P_{CO_2} were due both to the amount of calcite and to the size of the organic protoplast. The difference between Coulter counter versus flow cytometer volume measurements may be an effect of the different ways that the volume is calculated by these two instruments (electronically versus optically). nd, not determined.

	Coulter counter		Flow cytometer	
P_{CO_2}	Average coccosphere Coulter volume (μm^3)	Average coccolith volume (μm^3)	Average coccosphere side scatter (relative to 3-μm beads) (relative units)	Coccosphere forward-scatter volume before/after acidification (μm^3)	Average number of detached coccoliths per coccosphere
280.00	55.44	1.09	3.86	115/66.1	13.2
303.79	45.95	0.84	3.84	111/57.4	10.3
489.18	65.13	1.11	3.89	123/63.6	24.2
595.09	55.23	1.84	nd	nd	nd
750.25	69.33	1.86	4.05	155/77.1	80.3

Table 2. Carbonate chemistry in *E. huxleyi* cultures corresponding to different CO_2 scenarios from preindustrial time to projections for the end of this century (11). For each parameter, the numbers in the first row represent average values measured in the exponential growth phase, the numbers in the second row represent the blank values at the beginning of the experiment, and the values in the third row correspond to 1 SD of three samples.

Parameter	Preindustrial	Circa 1930	2035	2060	2100
P_{CO_2} (ppmv)	280.0	303.8	489.2	595.1	750.2
	268.2	326.3	524.8	726.2	844.1
	0.3	3.2	3.5	9.2	3.0
$[CO_2]$ (μmol liter^{-1})	9.5	10.2	16.4	19.9	25.1
	9.0	10.9	17.6	24.3	28.2
	0.0	0.0	0.1	0.3	0.1
$[CO_3^-]$ (μmol liter^{-1})	222.7	215.0	157.3	112.1	108.5
	244.0	216.4	157.7	123.3	110.0
	0.3	0.0	0.8	1.4	0.4
[DIC] (μmol liter^{-1})	1906.9	1923.5	2016.7	1848.1	2028.9
	1952.6	1993.2	2086.4	2136.1	2162.6
	0.7	0.6	1.0	1.1	0.2
$[HCO_3^-]$ (μmol liter^{-1})	1674.7	1698.2	1843.0	1716.0	1895.3
	1699.6	1765.8	1911.2	1988.4	2024.4
	0.3	0.6	1.7	2.2	0.4
Ω-calc	5.34	5.16	3.77	2.69	2.60
	5.85	5.19	3.78	2.96	2.64
	0.00	0.00	0.02	0.03	0.01
pH	8.15	8.13	7.96	7.85	7.79
	8.19	8.12	7.95	7.82	7.77
	0.00	0.00	0.00	0.01	0.00
Alkalinity (μeq liter^{-1})	2220.3	2224.9	2227.6	1995.9	2161.7
	2294.0	2292.8	2294.6	2288.3	2291.2
	1.2	0.6	0.3	1.3	0.4

tibilities to dissolution, indicates that our observed increase in coccolith mass cannot be accounted for by changing species compositions or dissolution effects (26).

The increase of ~4.5 pg in the average mass of $CaCO_3$ per coccolith since ~1960, as indicated by the smoothed least-squares curve in Fig. 4, coincides with rising atmospheric P_{CO_2} and is consistent in direction and relative magnitude with changes demonstrated here using laboratory experiments with *E. huxleyi* under future CO_2 scenarios. On average, 75% by mass of the <10 μm calcite (calculated by multiplying coccolith counts by typical coccolith volumes) at site RAPID 21-

12-B constitutes coccoliths of only two taxa, *C. pelagicus pelagicus* and *Calcidiscus leptoporus*, and just 3.2% comes from *E. huxleyi* (fig. S2). Typical coccoliths of the massive *C. pelagicus pelagicus* and *Calcidiscus leptoporus* species are approximately 15 and 7 times the average pre-1960 coccolith mass, respectively (27), and *C. pelagicus pelagicus* alone would require a <5% increase in coccolith mass (equivalent to a ~0.25-μm diameter increase) to account for the entire observed coccolith mass change, which is well within present-day variability (27). Therefore, because changes in the average coccolith mass can be dominated by only a small number of heavily calcifying spe-

cies, it is quite possible that the global calcification response may vary greatly with coccolithophore species assemblage in alternative ocean regimes. However, the dominance of *C. pelagicus pelagicus* over the sedimentary calcite mass observed in this core is typical within the North Atlantic (27–29), and therefore our findings probably represent a regional response, the response of a basin highly sensitive to anthropogenic CO_2 production (9). If species other than *E. huxleyi* exhibit a concomitant increase in PIC and POC production with rising CO_2 as demonstrated here for *E. huxleyi*, there would be no net change in this ratio with time, but we cannot quantify this ratio without a record of total organic carbon production. Nevertheless, a potential consequence of increasing calcification is a greater removal of POC from the surface waters because of increased ballast effects (30), although it is inconclusive whether or to what degree increased $CaCO_3$ ballast would favor a relative increase in POC export (31).

Discussion. Delving into the geological record potentially provides additional insight into coccolithophore response to elevated P_{CO_2}. Preservation of calcareous nannofossils relies on a buffering of the Ω-cal by vertical migration of the calcite compensation depth (CCD), the depth at which the rate of calcite input from surface waters equals the rate of dissolution. On time scales of >10,000 years, the CCD buffer keeps Ω-cal relatively constant (32); however, on shorter time scales there have been intervals in the geological past where the CCD has temporally shoaled, suggesting ocean acidification and transient decreases in carbonate saturation. The most widely studied of these intervals is the Paleocene Eocene Thermal Maximum (PETM, ~55 million years ago) (33). Calcareous nannofossil records suggest no obvious reduction in their abundance, shifts in distribution, or evolutionary bias attributable to ocean acidification during the PETM (34). The pH and P_{CO_2} reached in our culture experiments are within estimates of those indicated for the PETM (Fig. 1), and our laboratory and field results are again consistent with the lack of evidence for a change in saturation state being detrimental to coccolithophores.

Our single-species culture experiments and high-latitude assemblage records suggest that in a scenario where the P_{CO_2} in the world's oceans increases to 750 ppmv, coccolithophores will double their rate of calcification and photosynthesis (if ecosystem processes allow the survival of similar numbers of larger coccolithophore cells in the future). Given that coccolithophores are a major contributor [about 50% (35)] to the open-ocean carbonate pump, but a much smaller contributor [about 10% (36)] to the soft-tissue pump, we expect a disproportionate impact on overall community rates of calcification. Our experiments were conducted on *E. huxleyi*, which forms blooms at high latitudes that provide a snapshot of the response of *E. huxleyi* to P_{CO_2} under nutrient-replete conditions. Previous work using chemostat cultures under nutrient-limiting conditions (37) showed that increasing P_{CO_2} resulted in a decrease
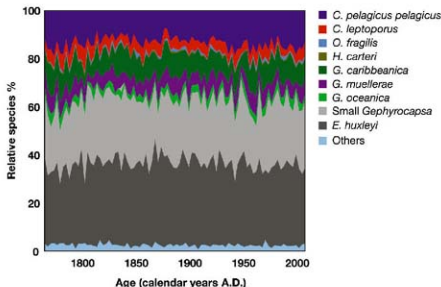

Fig. 3. Relative percentage abundance of coccoliths of each species in RAPID 21-12-B counted under a light microscope. No long-term trend in species composition was observed, indicating little or no species response to anthropogenic forcing. Stasis in the species composition, as would be expected considering the small temperature variation over this interval, implies that the core material is unaffected by dissolution (26), which was confirmed by SEM examination. The observed species assemblage is consistent with those published for other central subpolar Atlantic sites (27, 29).

Fig. 4. Average mass of $CaCO_3$ per coccolith in core RAPID 21-12-B and atmospheric CO_2. The average mass of $CaCO_3$ per coccolith in core RAPID 21-12-B (open circles) increased from 1.08×10^{-11} to 1.55×10^{-11} g between 1780 and the modern day, with an accelerated increase over recent decades. The increase in average coccolith mass correlates with rising atmospheric P_{CO_2}, as recorded in the Siple ice core (gray circles) (26) and instrumentally at Mauna Loa (black circles) (38), every 10th and 5th data point shown, respectively. Error bars represent 1 SD as calculated from replicate analyses. Samples with a standard deviation greater than 0.05 were discarded. The smoothed curve for the average coccolith mass was calculated using a 20% locally weighted least-squares error method.
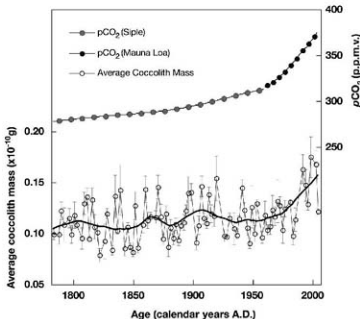
in net calcification rate and gross community production but had no noticeable effect on the ratio of calcification to photosynthesis. Other species need to be investigated in light of the variability encountered in response to changing P_{CO_2} between coccolithophore species that are representative of low and mid-latitudes (25).

Future research is needed to fully constrain productivity changes over the Anthropocene period, extend our understanding of calcification changes at different latitudes and in different ocean basins, and quantify how changing ballast will affect export production. The widely held assumption that all coccolithophores will decrease their calcification under elevated P_{CO_2} needs reappraisal in the light of our laboratory and field observations that demonstrate enhanced PIC production and cell size under high P_{CO_2} conditions and the resilience of calcifying phytoplankton in the geological record (34). Our analyses are highly relevant to ocean biogeochemical modeling studies and underline the physiological and ecological versatility of coccolithophores and their evolutionary adaptation through changes in ocean carbonate chemistry associated with past and projected P_{CO_2} levels.

References and Notes

1. E. Paasche, *Phycologia* 49, 503 (2002).
2. J. A. Kleypas *et al.*, *Science* 284, 118 (1999).
3. J. Bijma, H. J. Spero, D. W. Lea, B. E. Bemis, in *Use of Proxies in Paleoceanography: Examples from the South Atlantic*, G. Fischer, G. Wefer, Eds. (Springer, Berlin, 1999), pp. 489–512.
4. U. Riebesell *et al.*, *Nature* 407, 364 (2000).
5. I. Zondervan, R. E. Zeebe, B. Rost, U. Riebesell, *Global Biogeochem. Cycles* 15, 507 (2001).
6. B. Delille *et al.*, *Global Biogeochem. Cycles* 19, GB2023, 10.1029/2004GB002318 (2005).
7. G. Langer *et al.*, *Geochem. Geophys. Geosyst.* 7, Q09006 (2006).
8. S. Trimborn, G. Langer, B. Rost, *Limnol. Oceanogr.* 52, 2285 (2007).
9. R. A. Feely *et al.*, *Science* 305, 362 (2004).
10. K. P. Boessenkool, I. R. Hall, H. Elderfield, I. Yashayaev, *Geophys. Res. Lett.* 34, 10.1029/2007GL030285 (2007).
11. Materials and methods are available as supporting material on *Science* Online.
12. K. Caldeira, M. E. Wickett, *Nature* 425, 365 (2003).
13. J. Bollmann, J. O. Herrle, *Earth Planet. Sci. Lett.* 255, 273 (2007).
14. Z. S. Kolber, O. Prasil, P. G. Falkowski, *Biochim. Biophys. Acta* 1367, 88 (1998).
15. This "balance point" [the calcification-to-photosynthesis ratio at which a bloom has zero net impact on atmospheric CO_2 (C:P)] was calculated as having a value of about 1.5 at [DIC] = 2050 µmol kg^{-1}, [Alk] = 2300 µeq kg^{-1}, temperature = 15°C, and salinity = 35. The value of the balance point C:P ratio exhibits some variability with in situ conditions. For instance, it varies in the range 1.0 to 1.8 as in situ conditions vary in the ranges [DIC] = 2000 to 2300 µmol kg^{-1}, [Alk] = 2200 to 2400 µeq kg^{-1}, temperature = 10 to 30°C and salinity = 33 to 37. Its value will decrease by about 20% as P_{CO_2} increases from 280 ppmv to 700 ppmv (other conditions remaining constant) (16).
16. M. Frankignoulle, C. Canon, J.-P. Gattuso, *Limnol. Oceanogr.* 39, 456 (1994).
17. T. R. Anderson, P. Pondaven, *Deep-Sea Res. I* 50, 573 (2003).
18. U. Riebesell, I. Zondervan, U. Riebesell, *Limnol. Oceanogr.* 44, 683 (1999).
19. N. A. Nimer, M. J. Merrett, *New Phytol.* 123, 673 (1993).
20. P. Z. Wang, I. N. McCave, J. Geol. Soc. *Lond.* 147, 373 (1990).
21. M. Frenz, K. H. Baumann, B. Boeckel, R. Hoppner, R. Henrich, *J. Sediment. Res.* 75, 464 (2005).
22. I. N. McCave, I. R. Hall, *Geochem. Geophys. Geosyst.* 7, Q10N05 (2006).
23. J. A. Flores, F. J. Sierro, *Micropaleontology* 43, 321 (1997).
24. M. Frenz, R. Henrich, *Sedimentology* 54, 391 (2007).
25. L. Beaufort, I. Probert, N. Buchet, *Geochem. Geophys. Geosyst.* 8, Q09011, 10.1029/2006GC00149.
26. M. E. Hill, *Micropaleontology* 21, 227 (1975).
27. J. R. Young, P. Ziveri, *Deep-Sea Res. II* 47, 1679 (2000).
28. P. Ziveri, K. H. Baumann, B. Bockel, J. Bollmann, J. Young, in *Coccolithophores: From Molecular Processes to Global Impact*, H. Thierstein, J. Young, Eds. (Springer, Berlin, 2004), pp. 403–428.
29. A. McIntyre, A. W. H. Bé, *Deep-Sea Res.* 14, 561 (1967).
30. D. Archer, E. Maier-Reimer, *Nature* 367, 260 (1994).
31. S. Barker, J. A. Higgins, H. Elderfield, *Philos. Trans. R. Soc. London Ser. A* 361, 1977 (2003).
32. T. Tyrrell, R. E. Zeebe, *Geochim. Cosmochim. Acta* 68, 3521, 10.1016/j.gca.2004.02.018 (2004).
33. J. C. Zachos *et al.*, *Science* 308, 1611 (2005).
34. S. J. Gibbs, P. R. Bown, J. A. Sessa, T. J. Bralower, P. A. Wilson, *Science* 314, 1770 (2006).
35. K. H. Baumann, H. Andruleit, B. Bockel, M. Geisen, H. Kinkel, *Palaontol. Z.* 79, 93 (2005).
36. A. J. Poulton, T. R. Adey, W. M. Balch, P. M. Holligan, *Deep-Sea Res. II* 54, 538 (2007).
37. A. Sciandra *et al.*, *Mar. Ecol. Prog. Ser.* 261, 111 (2003).
38. L. J. Royer, R. A. Berner, I. P. Montañez, N. J. Tabor, D. J. Beerling, *GSA Today* 14, 10.1130/1052-5173(2004)0142.0.CO;2 (2004).
39. C. D. Keeling, T. P. Whorf, in *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, 2005).
40. J. D. L. van Bleijswik, R. S. Kempers, M. J. Veldhuis, *J. Phycol.* 30, 230 (1994).
41. The authors acknowledge M. Hill for manufacturing the bubbling flasks, B. Alker for her assistance in PIC preparation and analysis, D. Hydes for access to the Versatile Instrument for the Determination of Titration Alkalinity instrument, R. Head for POC analysis, and R. Gibson for assistance with PIC and FRRF analysis. We thank K. Medley for laboratory assistance and J. Elliot for useful discussions. We are grateful to the master, officers, crew, and scientific party of the RRS *Charles Darwin* cruise CD159, in particular I. N. McCave and H. Elderfield. We are grateful to O. M. Schofield, P. A. Tyler, and T. Anderson for discussions on the manuscript. We thank K. Davis for assistance with graphic illustrations. This work was supported by the Royal Society research grant no. 24437 (M.D.I.-R.) and by the Betty and Gordon Moore Foundation Marine Microbiology Investigator Award (flow cytometry analysis) (E.V.A.). The field work was supported by the UK Natural Environment Research Council RAPID Program, P.R.H. acknowledges support from Natural Environment Research Council (NERC) grant no. NER/S/S/2004/12772, R.E.M.R. and I.R.H. gratefully acknowledge NERC grant no. NER/T/S/2002/00980.

The Global Circulation of Seasonal Influenza A (H3N2) Viruses

Colin A. Russell,[1] Terry C. Jones,[1,2,3] Ian G. Barr,[4] Nancy J. Cox,[5] Rebecca J. Garten,[5] Vicky Gregory,[6] Ian D. Gust,[4] Alan W. Hampson,[4] Alan J. Hay,[6] Aeron C. Hurt,[4] Jan C. de Jong,[7] Anne Kelso,[4] Alexander I. Klimov,[5] Tsutomu Kageyama,[8] Naomi Komadina,[4] Alan S. Lapedes,[8] Yi P. Lin,[6] Ana Mosterin,[1,3] Masatsugu Obuchi,[7] Takato Odagiri,[7] Albert D. M. E. Osterhaus,[7] Guus F. Rimmelzwaan,[7] Michael W. Shaw,[5] Eugene Skepner,[1] Klaus Stohr,[9] Masato Tashiro,[7] Ron A. M. Fouchier,[2] Derek J. Smith[1,2*]

Antigenic and genetic analysis of the hemagglutinin of ~13,000 human influenza A (H3N2) viruses from six continents during 2002–2007 revealed that there was continuous circulation in east and Southeast Asia (E-SE Asia) via a region-wide network of temporally overlapping epidemics and that epidemics in the temperate regions were seeded from this network each year. Seed strains generally first reached Oceania, North America, and Europe, and later South America. This evidence suggests that once A (H3N2) viruses leave E-SE Asia, they are unlikely to contribute to long-term viral evolution. If the trends observed during this period are an accurate representation of overall patterns of spread, then the antigenic characteristics of A (H3N2) viruses outside E-SE Asia may be forecast each year based on surveillance within E-SE Asia, with consequent improvements to vaccine strain selection.

Influenza A (H3N2) virus is currently the major cause of human influenza morbidity and mortality worldwide. On average, influenza viruses infect 5 to 15% of the global population, resulting in ~500,000 deaths annually (1). Despite substantial progress in many areas of influenza research, questions such as when and to what extent the virus will change antigenically, and to what extent viruses spread globally, remain unanswered. A fundamental issue behind these questions is whether epidemics are the consequence of low-level persistence of viruses from the previous epidemic or whether they are seeded from epidemics in other regions and, if so, from where (2–8).

Addressing these issues of local persistence and global spread is vitally important for designing optimal surveillance and control strategies. If epidemics were regularly seeded from an outside region and if the source region of seed strains could be identified, it may be possible to forecast which variants would appear in epidemics in seeded

regions and, consequently, to optimize vaccine strain selection. Alternatively, if viruses persist within a region, evolve, and reemerge to cause the next epidemic, intervention strategies targeting virus circulation between epidemics might be effective in minimizing subsequent epidemics.

The eight gene segments of influenza A viruses reassort, leading to complicated phylogenetic patterns at the genomic scale (4, 8–10). The gene segment coding for the hemagglutinin (HA) is of major importance because the HA protein is the primary target of the protective immune response. Consequently, the HA is the focus of public health surveillance and the primary component of currently licensed influenza vaccines. We used antigenic and genetic analyses of the HA as a marker to investigate the global evolution and epidemiology of influenza A (H3N2) viruses from 2002 to 2007 and to determine whether influenza epidemics arise from locally persisting strains or whether epidemics are seeded from other regions.

Antigenic Evolution

Antigenic cartography has shown that the antigenic evolution of A (H3N2) virus, since its appearance in humans in 1968, can be represented by a two-dimensional (2D) antigenic map (11). Since 2002, the antigenic evolution of A (H3N2) viruses has roughly followed a line away from the A/Sydney/5/1997–like viruses that predominated in 1998 through the A/Fujian/441/2002–like viruses to the A/California/7/2004–like viruses to the A/Wisconsin/67/2005–like strains that predominated in 2006 and 2007. Figure 1A uses this directional bias to make a 2D plot to show spatial antigenic evolution and epidemiology in one plot.

Previously, Smith et al. (11) showed that, from 1968 to 2003, the antigenic evolution of A (H3N2) virus was punctuated: Periods of relative stasis lasting from 3 to 8 years were followed by rapid antigenic change, resulting in transitions to new antigenic clusters that necessitated an update of the influenza virus vaccine strain. During this time, several clusters also exhibited intracluster antigenic evolution, necessitating a within-cluster update of the vaccine strain. We find a general pattern of intracluster antigenic evolution from 2002 to 2007. During this time, antigenic evolution progressed away from A/Sydney/5/1997 at an average rate of 2.13 antigenic units per year (Fig. 1A). A distance of two antigenic units,

[1]Department of Zoology, University of Cambridge, Cambridge, UK. [2]Department of Virology, Erasmus Medical Centre, Rotterdam, Netherlands. [3]Universitat Pompeu Fabra, Barcelona, Spain. [4]World Health Organization (WHO) Collaborating Centre for Reference and Research on Influenza, Melbourne, Australia. [5]WHO Collaborating Center for Influenza, Centers for Disease Control and Prevention, Atlanta, GA, USA. [6]WHO Collaborating Centre for Influenza, National Institute for Medical Research (NIMR), London, UK. [7]WHO Collaborating Center for Influenza, National Institute for Infectious Diseases, Tokyo, Japan. [8]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. [9]Novartis Vaccines and Diagnostics, Cambridge, MA, USA.

*To whom correspondence should be addressed. E-mail: dsmith@zoo.cam.ac.uk

representing a fourfold difference in hemagglutination inhibition (HI) assay titer, is generally considered as a sufficient antigenic difference to warrant a vaccine update. The A (H3N2) component of the influenza vaccine was updated four times during this period. A core component of vaccine strain selection involves identifying emerging antigenic variants. If an emerging variant is judged likely to cause epidemics in the upcoming influenza season, the vaccine is updated to contain a representative of the new strain.

We find differences in the amount of antigenic variation seen during an epidemic in different regions and from year to year in the same region (Fig. 1A). Despite such spatial and temporal heterogeneities, the antigenic evolution has been markedly homogeneous on a global scale. An explanation for this homogeneity could be that viruses circulate globally rather than persist and evolve locally.

To search for global patterns in the source of emerging variants, we measured which regions were leading or trailing antigenically and found that, from 2002 to 2007, newly emerged strains of the A (H3N2) subtype appeared in E-SE Asian countries, on average, ~6 to 9 months earlier than they appeared in other regions, with long delays to South America, typically of an additional ~6 to 9 months (Fig. 2A). Though A (H3N2) viruses in E-SE Asian countries are on average more antigenically advanced, there is sufficient variability from season to season (colored circles in Fig. 2A) such that no one particular country in the region is consistently most advanced. Thailand, Malaysia, and Japan are exceptions to the Asia-leading pattern, being less antigenically advanced than the rest of the region.

One interpretation of this Asia-leading pattern is that new variants emerge first in E-SE Asia and subsequently seed other regions of the world. An alternative but more complex explanation is that this pattern is the product of independent local persistence in multiple regions and parallel evolution in which similar antigenic variants emerge independently worldwide as a result of similar selection pressures. To test between these two interpretations, we must answer the fundamental long-standing question of whether influenza viruses persist in a region, and could thus undergo parallel evolution, or whether regions are regularly seeded from external regions.

About 10% of these ~13,000 A (H3N2) viruses analyzed antigenically were also analyzed genetically by sequencing the HA1 domain of the hemagglutinin (12). This subset was a representative sample of all ~13,000 isolates (fig. S1B) and thus was suitable for investigating the ancestry of strains and the fundamental issue of local persistence versus seeding in the global circulation of A (H3N2) viruses. The genetic progression over time (Fig. 1B) shows similar average patterns to the antigenic data, with Asia leading [as previously shown for Taiwan (13)] and South America trailing (Fig. 2, A and B). However, there were important differences (China in 2005 and Oceania in 2005). For influenza vaccine strain selection, genetic-antigenic

differences are resolved in favor of the antigenic data, because the humoral immune system "sees" the virus phenotype, not the genotype.

Persistence Versus Seeding

Source of inter-epidemic strains. The simplest test of persistence versus seeding is to examine the origin of strains isolated between epidemics. If viruses persist locally, at least some of the inter-epidemic strains would be descended from, and thus more closely related to, strains from the previous epidemic than to strains from outside the region. Alternatively, if there was no persistence, inter-epidemic strains would be more similar to strains from elsewhere.

We sequenced the HA1 domains of 52 inter-epidemic strains isolated in Oceania (primarily Australia and New Zealand), North America, and Japan from June 2002 to September 2006. None of these inter-epidemic strains was more similar to strains from the previous local epidemic than to externally circulating strains (Fig. 3). This result is evidence for external seeding and against local persistence.

Even when done well, inter-epidemic surveillance yields relatively small amounts of data that can never completely rule out the existence of local virus persistence between epidemics, especially of any low-pathogenic variants that produce subclinical infections. The test described in the next section has the advantage of using all available sequence data rather than being limited to inter-epidemic data. Nevertheless, all tests must take into account the effects of external introductions during local epidemics (14).

Evolutionary relationship of strains from one epidemic to the next in a region. As described by Nelson et al. (4, 8), if epidemic strains persist locally and give rise to the next local epidemic, those strains should be more closely related to one another than to strains isolated in other regions, and a phylogenetic tree of the data would look like that depicted in fig. S2A. Conversely, if epidemics were seeded from outside a region, the epidemic strains would be more similar to contemporary strains from outside that region than to strains from the previous local epidemic (fig. S2B).

Following the methodology of Nelson et al. (4, 8), we constructed a phylogenetic tree of the HA1 domain of the hemagglutinin from the sequenced subset of the global surveillance data (Fig. 1C and fig. S3A). In this tree, the HA1 of the viruses in each epidemic in a temperate region (four in North America, five in Oceania, four in Europe, and four in Japan) and each epidemic in a subtropical region (three in Hong Kong) descended from externally circulating strains, not from strains in the previous local epidemic (Fig. 1, C and D). The topology of this tree is more similar to that in fig. S2B than to that in fig. S2A. This result is also evidence for external seeding and against persistence. For other regions, including most tropical and subtropical regions, there were fewer sequences, and it was not possible to conclusively differentiate between persistence and seeding.
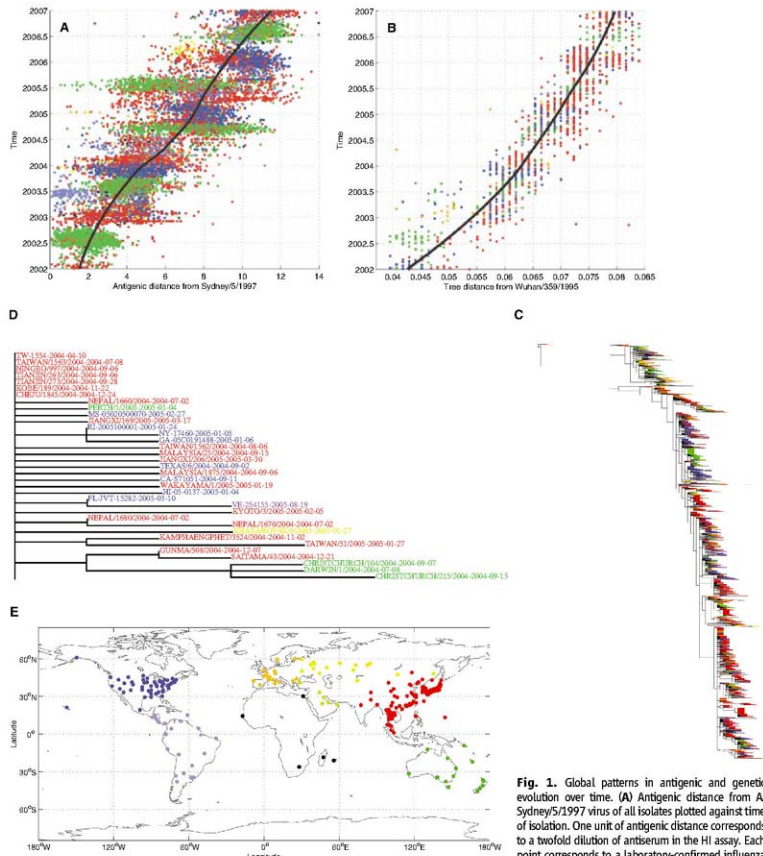
Fig. 1. Global patterns in antigenic and genetic evolution over time. (**A**) Antigenic distance from A/Sydney/5/1997 virus of all isolates plotted against time of isolation. One unit of antigenic distance corresponds to a twofold dilution of antiserum in the HI assay. Each point corresponds to a laboratory-confirmed influenza A (H3N2) isolate, with the color of the point indicating the region of geographic origin as shown in (E). The thick black line is the best-fit statistical model (a loess spline) fitted through all data points to show the trend over time. Points in advance of the spline are antigenically advanced, whereas strains behind the spline are antigenically lagging. (**B**) Genetic version of (A) for all sequenced strains with distance measured to the root, A/Wuhan/359/1995, of a maximum likelihood nucleotide phylogenetic tree (C). The thick black line is the same as in (A) but for the genetic data. (**C**) Phylogenetic tree of HA1 nucleotide sequences color-coded by geographic origin (E), including strain names and isolation dates. We constructed the initial tree with the PhyML software package version 2.4.5 (38), with 1567 nucleotide sequences (12) and A/Wuhan/359/1995 as the root, using GTR+I+Γ₄ (the general time-reversible model with the proportion of invariant sites and the gamma distribution of among-site rate variation with four rate categories estimated from the empirical data) [determined by ModelTest (39)] as the evolutionary model. GARLI version 0.951 (40) was run on the best tree from PhyML for two million generations to optimize tree topology and branch lengths. Figure S3A provides a "zoom-able" version of this image. (**D**) Partial detail of (C). (**E**) Color-coded geographic setting for (A) to (D).

This evidence for external seeding and against persistence agrees with full-genome analyses of New York state, Australia, and New Zealand data that show global migration of A (H3N2) viruses rather than local persistence (4, 8–10, 15). In addition, Nelson *et al.* (8) find evidence compatible with either northern-to-southern hemisphere migration or migration from tropical regions, including Southeast Asia.

Source of Seeding

Given this evidence for seeding and against local persistence, we are left with the interpretation that the E-SE Asia–leading pattern (Fig. 2, A and B) implies that new variants emerge first in E-SE Asia and then seed the rest of the world. The phylogenetic tree provides further evidence to support this interpretation, showing that the ancestors of strains in temperate regions typically originate in E-SE Asia, with the "trunk" of the phylogenetic tree typically occupied by E-SE Asian strains (Fig. 1, C and D) and with, on average, E-SE Asian strains closer to the trunk ($P < 0.001$) than strains from all other regions (Fig. 2C).

The above analyses, in addition to being evidence for an "out of E-SE Asia" hypothesis, are also evidence against several other long-standing unresolved hypotheses for the global circulation of influenza viruses, as follows.

Out-of-China. If China alone were the source of all new variants and effectively seeded the rest of the world (16), then Chinese strains would be (i) closer to the trunk than strains from all other regions each year and (ii) consistently antigenically and genetically advanced relative to strains from other regions, both of which we do not find (Fig. 2). Commensurate with previously published studies, we find new variants are sometimes first detected in China (15, 17, 18); however, we also find that new variants are sometimes first detected in other countries in E-SE Asia (Fig. 2).

Out-of-tropics. In this hypothesis, epidemics in regions outside the tropics are seeded from the tropics (7, 19). If true, A (H3N2) viruses would need to circulate continually in the tropics and would give rise to one of two evolutionary patterns. One pattern would arise if tropical Asia, Africa, and South America were well connected epidemiologically to one another; in this case, all three regions would be similarly antigenically and genetically advanced and closer to the trunk than nontropical regions. The other pattern would arise if the three tropical regions were poorly connected epidemiologically to one another; in this case, there would be independent genetic lineages for each tropical region. There is currently not enough data to include tropical Africa in this analysis. However, for tropical Asia and South America, neither of these patterns is observed (Figs. 1C and 2 and fig. S3A) (14).

Seeding by strains moving between the northern and southern hemispheres. If this hypothesis were true, then each year epidemics in the northern hemisphere would be seeded by viruses from epidemics in the southern hemisphere and vice versa (20–22). The phylogenetic tree for the HA1 domain (Fig. 1C and fig. S3A) shows no evidence for epidemic strains in the northern hemisphere being descendant from strains epidemic in South America or Africa. The tree shows limited evidence for Oceania playing a role in seeding a minority of epidemics in the northern hemisphere (14) but at a level insufficient to support this hypothesis as the dominant mechanism of the global circulation of influenza viruses. This hypothesis also fails to explain how viruses from Asia, which is almost entirely contained within the northern hemisphere, lead antigenically and genetically and are closest to the trunk of the phylogenetic tree.

Local persistence with seeding only at cluster transitions. Though there was no major cluster transition during our study period, there has been an average of 2.13 units of antigenic evolution each year. Even so, in temperate regions, no local persistence could be detected, and each epidemic was seeded by exogenous strains. The drift variants observed in this study have emerged from E-SE Asia; however, because we have not seen a major cluster transition of the magnitude of Wuhan 1995 to Sydney 1997 (~4.7 antigenic units), we can neither exclude that in such a case a new variant could emerge outside E-SE Asia nor that it could affect seeding patterns.

The E-SE Asian Circulation Network

For E-SE Asia to seed epidemics in multiple regions of the world, influenza virus must circulate continually in E-SE Asia. But how?
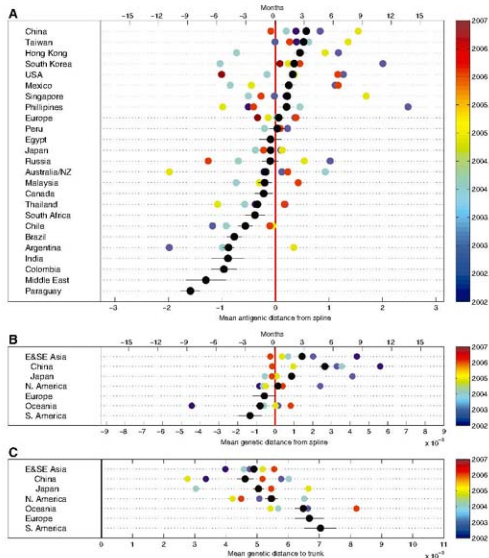

Fig. 2. Evolutionary leading and trailing regions. (**A**) Black circles indicate the average antigenic distance to the spline of Fig. 1A for all strains isolated in a region, and the thin horizontal black line indicates the SEM. Colored circles split this overall average by epidemic; circle color indicates time. The spline can also be interpreted as a function of time; thus, time is shown as a second x axis. (**B**) Similar to (A) but based on genetic distance to spline from Fig. 1B. (**C**) Genetic distance to trunk of the phylogenetic tree by region and season. We algorithmically defined the trunk of the tree in Fig. 1C (14) and calculated the tree distance of each strain to the trunk. Average distance to trunk was calculated per region and per season. The black circles indicate the overall average per region, the thin horizontal black line indicates the SEM, and colored circles indicate seasonal averages. The mean for E-SE Asia is different from that of Oceania ($P < 0.00001$), North America ($P < 0.001$), Europe ($P < 0.01$), and South America ($P < 0.0001$).

It is generally considered that influenza viruses continually circulate in tropical countries (7, 23–27) and, if this were true, it would explain how influenza viruses could persist in tropical Asia. Indeed, circulation in an endemic core area that seeds satellite areas has been shown to be a key epidemiological process for the continual circulation of antigenically stable pathogens (28, 29). Reports based on influenza-like illness (ILI) or influenza and pneumonia mortality (IPM) data describe continual circulation in the tropics (23). However, several viruses other than influenza can cause ILI and IPM, and studies from tropical countries based on viral isolations show a marked seasonality for influenza epidemics, with peaks usually occurring during periods of high rainfall (30–35). In agreement with these studies based on virus isolation, our virus isolation study also finds that influenza has clear epidemic peaks and deep troughs in all regions, including the four tropical and four subtropical E-SE Asian countries for which there are sufficient data to detect an epidemic signal. Thus, continuous circulation in individual tropical countries is unlikely to be the mechanism for persistence in E-SE Asia. However, more data from a wide diversity of locations are needed to fully understand seasonal forcing and to definitively exclude local persistence as an element of transmission dynamics in tropical and subtropical areas of E-SE Asia.

Another possibility for continual circulation is that viruses pass from epidemic to epidemic among countries via the mobile human population. Figure 4, A and C, shows that there is sufficient variability in the timing of epidemics within E-SE Asia such that the virus could circulate continuously in this way as a result of the temporal overlap of epidemics. Much of the variability in the timing of epidemics is likely to be linked to the heterogeneity in the timing of lower

Fig. 3. The genetic relationship of interseasonal strains to strains in the previous local epidemic and to strains epidemic in other regions. Interseasonal strains are defined as strains isolated more than 1 month after the end of the previous local epidemic and more than 1 month before the beginning of the next local epidemic. For each interseasonal strain, the phylogenetic tree distance was calculated to the closest strain in the previous local epidemic and to the closest strain found outside the region in the previous 4 months. The diagonal line is 1:1 and is included for reference.

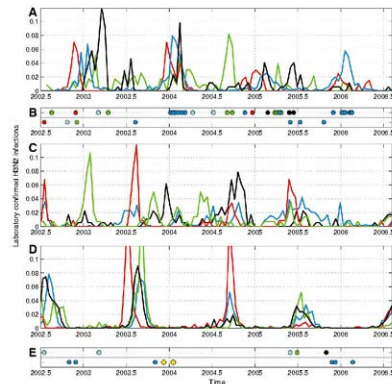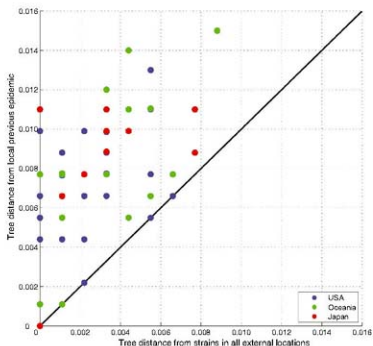**Fig. 4.** Synchrony of epidemics in east Asia and the South Pacific. (**A**) Epidemics in east Asia. The *y* axis shows laboratory-confirmed H3N2 infections per 2 weeks as a proportion of the total number of laboratory-confirmed H3N2 infections over the study period in each location. (**B**) Strains on the trunk of the phylogenetic tree are of particular evolutionary importance in testing for virus migration among countries. In (B) and (E), there is a circle for every strain on the trunk of the phylogenetic tree (figs. S3A and S4). The purpose is to show where the trunk strains were isolated [top row color code from (A), bottom row from (C)] and when they were isolated, to assess the epidemiological activity at the time of isolation. Cyan circles represent E-SE Asian strains but in locations not shown in (F). (**C**) Same as (A) but for tropical Southeast Asia. (**D**) Same as (A) but for Australia and New Zealand. (**E**) Same as (B), but the top row are Oceanian strains [cyan circles represent strains from cities in Australia not shown in (F)], and the bottom row are strains from North America (blue) and Russia and Ukraine (yellow). (**F**) Geographic setting for (A) to (E).

temperatures and rainy seasons (19, 33, 36). We thus hypothesize that the variability of epidemics, combined with the interconnectedness of E-SE Asian countries, forms an E-SE Asian circulation network that maintains influenza virus in the region by passing from epidemic to epidemic.

If such a network existed, we would expect a temporal and phylogenetic progression of E-SE Asian viruses on the trunk of phylogenetic tree as viruses pass from epidemic to epidemic within the network. Figure S4 shows such a progression, and Fig. 4, B and E, shows the relationship of these trunk strains to the timing of epidemics. Trunk strains were isolated in temperate, subtropical, and tropical regions of E-SE Asia, indicating that all three climatic regions of E-SE Asia are part of the circulation network. To test whether the non–E-SE Asian strains on the trunk indicate that the circulation network includes countries outside of E-SE Asia or whether they represent one-way seeding events out of E-SE Asia, we examined the phylogenetic tree (Fig. 1C and fig. S3A). We found only limited instances of such seeding back into E-SE Asia, with clear evidence that most E-SE Asian strains were directly descendent from other E-SE Asian strains (14)—thus indicating that the temporally overlapping epidemics in E-SE Asia form a circulation network that, during the study period, has been mostly closed to external reseeding.

E-SE Asia's strong travel and trade connections to Oceania, North America, and Europe (14, 37) facilitate the rapid movement of new influenza virus variants into those areas and thus explain the relatively small lag in antigenic and genetic advancement seen in those regions (Fig. 2, A and B). Also, though it is unclear how much travel there must be between two locations for them to be epidemiologically well-connected, South America's 6- to 9-month antigenic lag (Fig. 2A) may be attributable to its paucity of direct connections with E-SE Asia (fig. S5). South America's strong travel connections to Europe and North America, but not to E-SE Asia, could result in a seeding hierarchy where strains are first seeded into North America and Europe and from there to South America (Fig. 5). Most strains appear to circulate in this simple hierarchy, and even those strains that circulate in a more complex hierarchy still originate in E-SE Asia (14). Thus, the extinction of many H3 lineages—a key characteristic of the H3 phylogeny—may, in addition to the accumulation of deleterious mutations (25), also be due to reaching the end of this hierarchy.

Surveillance and Vaccine Strain Selection

A major practical function of WHO's Global Influenza Surveillance Network is to assist regulatory authorities to recommend which strains should be included in influenza vaccines. Expanding surveillance within the E-SE Asian circulation network will aid the early detection of the emergence and spread of new variants and could help to more precisely define the network. Such surveillance is crucial for optimizing vaccine strain selection for countries within the network and for forecasting which variants will seed epidemics in the rest of the world, consequently increasing vaccine efficacy and ultimately reducing influenza morbidity and mortality worldwide.

Given the importance of the HA, it is the only portion of the virus genome that is currently sequenced routinely within the WHO Global Influenza Surveillance Network. Recently, whole-genome sequencing initiatives have provided important insights into the genesis and spread of reassortment viruses, their rapid migration, and the cocirculation of multiple lineages (4, 8, 9, 15). Expanded sequencing of whole genomes will provide additional markers for tracking the global migration of viruses and reveal potential differences between the global evolution of the HA and the other gene segments. Such sequencing efforts should include strains from E-SE Asia and be linked with antigenic data on HA and, in the longer term, with phenotype changes determined by other virus genes to fully understand the selection pressures on influenza viruses and their epidemiology.

The data used in this study were generated by the WHO Global Influenza Surveillance Network. Although there are biases in surveillance data, these biases do not have a substantial effect on the results (14). The methods we have used are generic and, although applied here to human influenza A (H3N2) viruses, are broadly applicable to influenza viruses in other species and to other pathogens.

Summary

We present evidence from antigenic and genetic analyses of HA that, from 2002 to 2007, influenza A (H3N2) virus epidemics worldwide were seeded each year by viruses that originated in E-SE Asia. We find evidence that temporally overlapping epidemics in E-SE Asia create a circulation network in which influenza A (H3N2) viruses continually circulate within the region by passing from epidemic to epidemic. E-SE Asia's strong travel and trade connections with Oceania, North America, and Europe, coupled with weak connections to South America, could explain the seeding hierarchy observed in this period where new virus variants first seed epidemics in Oceania, North America, and Europe and later in South America. The mostly one-way nature of this hierarchy suggests that, once A (H3N2) viruses leave E-SE Asia, they are unlikely to contribute to long-term viral evolution. If the trends observed during this period are an accurate representation of overall patterns of spread, then the antigenic characteristics of A (H3N2) viruses outside E-SE Asia may be forecast each year based on surveillance within E-SE Asia, with consequent improvements to vaccine strain selection and reductions in influenza A (H3N2) morbidity and mortality. Intensified surveillance, including whole-genome sequencing, and better understanding of the evolutionary selection pressures in E-SE Asia would further improve vaccine strain selection worldwide and potentially make influenza virus evolution more predictable.

References and Notes

1. K. Stohr, *Lancet Infect. Dis.* **2**, 517 (2002).
2. Commission on Acute Respiratory Diseases, *Am. J. Hyg.* **47**, 290 (1948).
3. J. A. Dudgeon, C. H. Stuart-Harris, C. H. Andrewes, R. E. Glover, W. H. Bradley, *Lancet* **248**, 627 (1946).
4. M. I. Nelson et al., *PLoS Pathog.* **2**, e125 (2006).
5. R. M. Taylor, *Am. J. Public Health* **39**, 171 (1949).
6. S. B. Thacker, *Epidemiol. Rev.* **8**, 129 (1986).
7. C. Viboud, W. J. Alonso, L. Simonsen, *PLoS Med.* **3**, e89 (2006).
8. M. I. Nelson, L. Simonsen, C. Viboud, M. A. Miller, E. C. Holmes, *PLoS Pathog.* **3**, e131 (2007).
9. E. C. Holmes et al., *PLoS Biol.* **3**, e300 (2005).
10. E. Ghedin et al., *Nature* **437**, 1162 (2005).
11. D. J. Smith et al., *Science* **305**, 371 (2004).
12. Sequences were deposited in the GenBank database (accession numbers EU501119 to EU502618 and EU514616 to EU514683). Table S1 contains the sequence name, accession number, and country of isolation for each strain.
13. S. R. Shih et al., *J. Clin. Microbiol.* **43**, 1651 (2005).
14. Materials and methods are available as supporting material on *Science* Online.
15. L. Simonsen et al., *Mol. Biol. Evol.* **24**, 1811 (2007).
16. K. F. Shortridge, *Chin. Med. J. (Engl.)* **110**, 637 (1997).
17. N. J. Cox, T. L. Brammer, H. L. Regnery, *Eur. J. Epidemiol.* **10**, 467 (1994).
18. N. J. Cox, K. Subbarao, *Annu. Rev. Med.* **51**, 407 (2000).
19. W. J. Alonso et al., *Am. J. Epidemiol.* **165**, 1434 (2007).
20. W. P. Glezen, R. B. Couch, in *Viral Infections of Humans*, A. S. Evans, Ed. (Plenum Medical Book Company, New York, 1989), pp. 419–449.
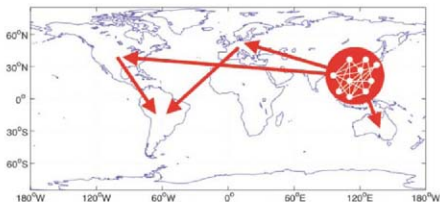

Fig. 5. Schematic of the dominant seeding hierarchy of seasonal influenza A (H3N2) viruses. The structure of the network within E-SE Asia is unknown.

21. J. S. Nguyen-Van-Tam, in *Textbook of Influenza*, K. G. Nicholson, R. G. Webster, A. J. Hay, Eds. (Blackwell Science, London, 1998), pp. 181–206.
22. C. Viboud *et al.*, *Emerg. Infect. Dis.* **10**, 32 (2004).
23. K. A. Fitzner, S. M. McGhee, A. J. Hedley, K. F. Shortridge, *Hong Kong Med. J.* **5**, 87 (1999).
24. A. S. Monto, *Int. Congr. Ser.* **1263**, 3 (2004).
25. M. I. Nelson, E. C. Holmes, *Nat. Rev. Genet.* **8**, 196 (2007).
26. J. M. Simmerman *et al.*, *Vaccine* **23**, 182 (2004).
27. C. M. Wong *et al.*, *PLoS Med.* **3**, e121 (2006).
28. B. T. Grenfell, O. N. Bjornstad, J. Kappey, *Nature* **414**, 716 (2001).
29. B. T. Grenfell, B. M. Bolker, *Ecol. Lett.* **1**, 63 (1998).
30. F. T. Chew, S. Doraisingham, A. E. Ling, G. Kumarasinghe, B. W. Lee, *Epidemiol. Infect.* **121**, 121 (1998).
31. T. S. David-West, A. R. Cooke, *J. Med.* **103** (1974).
32. E. de Arruda *et al.*, *J. Infect. Dis.* **164**, 252 (1991).
33. S. Doraisingham, K. T. Goh, A. E. Ling, M. Yu, *Bull. WHO* **66**, 57 (1988).
34. A. Dosseh, K. Ndiaye, A. Spiegel, M. Sagna, C. Mathiot, *Am. J. Trop. Med. Hyg.* **62**, 639 (2000).
35. B. L. Rao, K. Banerjee, *Bull. WHO* **71**, 177 (1993).
36. L. P.-C. Shek, S.-W. Lee, *Paediatr. Respir. Rev.* **4**, 105 (2003).
37. L. Hannaguel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15124 (2004).
38. S. Gründon, O. Gascuel, *Syst. Biol.* **52**, 696 (2003).
39. D. Posada, K. A. Crandall, *Bioinformatics* **14**, 817 (1998).
40. D. J. Zwickl, thesis, University of Texas (2006).
41. We thank the many individuals throughout the WHO Global Influenza Surveillance Network, particularly those in National Influenza Centers, for their enormous contributions to surveillance; the maintainers of the Influenza Sequence Database (www.flu.lanl.gov), D. Burke, D. Horton, G. Lewis, N. Lewis, B. Mansell, D. M. Smith, K. Waichert, and J.-C. Yeh; and the reviewers whose comments and suggestions substantially improved the manuscript. This work was supported by an NIH Director's Pioneer Award to D.J.S., part of the NIH roadmap for medical research, through grant number DP1-OD000490-01. R.A.M.F. is supported by National Institute of Allergy and Infectious Diseases–NIH contract HHSN266200700010C, as well as by the De Nederlandse Organisatie voor Wetenschappelijk Onderzoek Netherlands Influenza Research Centre. The Melbourne WHO Collaborating Centre for Reference and Research on Influenza is supported by the Australian Government Department of Health and Aging. The WHO Collaborating Centre for Influenza, NIMR, UK, is funded by The Medical Research Council (UK). For the antigenic cartography software, see www.antigenic-cartography.org. The conclusions presented in this paper are those of the authors and do not necessarily reflect those of the funding agencies.

REPORTS

Generalized Voice-Leading Spaces

Clifton Callender,[1] Ian Quinn,[2] Dmitri Tymoczko[3*]

Western musicians traditionally classify pitch sequences by disregarding the effects of five musical transformations: octave shift, permutation, transposition, inversion, and cardinality change. We model this process mathematically, showing that it produces 32 equivalence relations on chords, 243 equivalence relations on chord sequences, and 32 families of geometrical quotient spaces, in which both chords and chord sequences are represented. This model reveals connections between music-theoretical concepts, yields new analytical tools, unifies existing geometrical representations, and suggests a way to understand similarity between chord types.

To interpret music is to ignore information. A capable musician can understand the sequence of notes (C_4, E_4, G_4) in various ways: as an ordered pitch sequence (for example, an ascending C-major arpeggio starting on middle C), an unordered collection of octave-free note-types (for example, a C major chord), an unordered collection of octave-free note-types modulo transposition (for example, a major chord), and so on. Musicians commonly abstract away from five types of information: the octave in which notes appear, their order, their specific pitch level, whether a sequence appears right-side up or upside down (inverted), and the number of times a note appears. Different purposes require different information; consequently, there is no one optimal degree of abstraction.

Here we model this process. We represent pitches by the logarithms of their fundamental frequencies, setting middle C to 60 and the octave equal to 12. A musical object is a sequence of pitches ordered in time or by instrument (I). The object (C_4, E_4, G_4) can represent consecutive pitches played by a single instrument or a simultaneous event in which the first instrument plays C_4, the second E_4, and the third G_4. (Instruments can be ordered arbitrarily.) Musicians generate equivalence classes (2, 3) of objects by ignoring five kinds of transformation: octave shifts (O), which move any note in an object into any other octave; permutations (P), which reorder an object; transpositions (T), which move all the notes in an object in the same direction by the same amount; inversions (I), which turn an object upside down; and cardinality changes (C), which insert duplications into an object (4) (figs. S1 and Table 1). (Note that O operations can move just one of an object's notes, whereas T operations

move all notes.) We can form equivalence relations with any combination of the OPTIC operations, yielding $2^5 = 32$ possibilities.

A musical progression is an ordered sequence of musical objects. Let $\mathcal{F}$ be a collection of musical transformations, with $f, f_1,...,f_n \in \mathcal{F}$. The progression ($p_1$, ..., p_n) is uniformly $\mathcal{F}$-equivalent to [$f(p_1)$, ..., $f(p_n)$] and individually $\mathcal{F}$-equivalent to [$f_1(p_1)$, ..., $f_n(p_n)$]. Uniform equivalence uses a single operation to transform each object in the first progression into the corresponding object in the second; individual equivalence may apply different operations to a progression's objects (fig. S2). The OPTIC operations can be applied individually, uniformly, or not at all, yielding $3^5 = 243$ equivalence relations on progressions.

A number of traditional music-theoretical concepts can be understood in this way, including chord (OPC), chord type (OPTC), set class (OPTIC), chord-progression (individual OPC), voice leading (uniform OP), pitch class (single notes under O), and many others [table S1 and (4)]. We can also combine OPTIC operations in new ways, producing new music-theoretical tools. For example, analogs to voice leadings

Table 1. Equivalence relations and quotient spaces produced by the five principal transformations in Western music theory. Here, x is a point in $\mathbb{R}^n$, $\mathbf{1}$ represents $(1, ..., 1)$, and $\mathcal{S}_n$ is the symmetric group of order n.

	Equivalence relation	Space
None		$\mathbb{R}^n$
Octave	$x \sim_O x + 12i$, $i \in \mathbb{Z}^n$	$\mathbb{T}^n$
Transposition	$x \sim_T x + c\mathbf{1}$, $c \in \mathbb{R}$	$\mathbb{R}^{n-1}$ or $\mathbb{T}^{n-1}$ (if in conjunction with O) (orthogonal projection creates a barycentric coordinate system) and $/S_n$
Permutation	$x \sim_P \sigma(x)$, $\sigma \in \mathcal{S}_n$	
Inversion	$x \sim_I -x$	Add $/\mathbb{Z}_2$ [or $/(\mathcal{S}_n \times \mathbb{Z}_2)$ in conjunction with P]
Cardinality	$(..., x_i, x_{i+1} ...) \sim_C$ $(..., x_i, x_i, x_{i+1} ...)$	Infinite dimensional "Ran space"

[1]College of Music, Florida State University, Tallahassee, FL 32306, USA. [2]Music Department, Yale University, New Haven, CT 06520, USA. [3]Music Department, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: dmitri@princeton.edu

connect the elements of one chord type (or set class) to those of another; these are OPT (or OPTI) voice-leading classes, resulting from the application of uniform OP (or OPI) and individual T (1, 5) (Fig. 1). These equivalence relations can reveal connections within and across musical works and can simplify the analysis of voice leading by grouping the large number of possibilities into more manageable categories.

Geometrically, a musical object can be represented as a point in $\mathbb{R}^n$. The four OPTI equivalences create quotient spaces by identifying (or "gluing together") points in $\mathbb{R}^n$ (fig. S3). Octave equivalence identifies pitches p and $p +$ 12, transforming $\mathbb{R}^n$ into the n-torus $\mathbb{T}^n$. Transpositional equivalence identifies points in $\mathbb{R}^n$ with their (Euclidean) orthogonal projections onto the hyperplane containing chords summing

to 0. This transforms $\mathbb{R}^n$ into $\mathbb{R}^{n-1}$, creating a barycentric coordinate system in the quotient (basis vectors pointing from the barycenter of a regular n-simplex to its vertices). Permutation equivalence identifies points in $\mathbb{R}^n$ with their reflections in the hyperplanes containing chords with duplicate notes. Musical inversion is represented by geometric inversion through the origin. Permutation and inversion create singular quotient spaces (orbifolds) not locally Euclidean at their fixed points. C equivalence associates points in spaces of different dimension: The result is the infinite-dimensional union of a series of finite subset spaces (6–8).

One can apply any combination of the OPTI equivalences to $\mathbb{R}^n$, yielding $2^4 = 16$ quotient spaces for each dimension (Table 1); applying C produces 16 additional infinite-dimensional quo-

tients. Any ordered pair of points in any quotient space represents an equivalence class of progressions related individually by the relevant combination of OPTIC equivalences. The image of a line segment in $\mathbb{R}^n$ [a "line segment" in the quotient, although it may "bounce off" a singularity (1, 9)] can be identified with an equivalence class of progressions related uniformly by the relevant combination of OPIC and individually by T. (This is because T acts by orthogonal projection.) Intuitively, pairs of points represent successions between equivalence classes, considered as indivisible harmonic wholes; line segments represent specific connections between their elements.

Music theorists have proposed numerous geometrical models of musical structure (fig. S4 and table S3), many of which are regions of the spaces described in this report. These models have often been incomplete, displaying only a portion of the available chords or chord types and omitting singularities and other nontrivial geometrical and topological features. Furthermore, they have been explored in isolation, without an explanation of how they are derived or how they relate (10). Our model resolves these issues by describing the complete family of continuous n-note spaces corresponding to the 32 OPTIC equivalence relations.

Of these, the most useful are the OP, OPT, and OPTI spaces, representing voice-leading relations among chords, chord types, and set classes, respectively (4). The OP spaces $\mathbb{T}^n/\mathcal{S}_n$ (n-tori modulo the symmetric group) have been described previously (9). The OPT space $\mathbb{T}^{n-1}/\mathcal{S}_n$ is the quotient of an $(n-1)$-simplex, whose boundary is singular, by the rigid transformation cyclically permuting its vertices (4). The OPTI space $\mathbb{T}^{n-1}/(\mathcal{S}_n \times \mathbb{Z}_2)$ is the quotient of the resulting space by an additional

Fig. 1. Progressions belonging to the same OPT and OPTI voice-leading classes. Each group exhibits the same underlying voice-leading structure: Analogous elements in the first chord are connected to analogous elements in the second, and the distances moved by the voices are equal up to an additive constant. (A) A iv⁶-V⁷ progression from Mozart's C minor fantasy, Köchel catalog number (K.) 457, measures 13 and 14. (B) A progression from mm. 15-16 of the same piece, individually T-related to (A). (C) A progression from Beethoven's Ninth Symphony, movement I, measure 102, related to (A) by individual T and uniform OPI. (D) A common voice leading between fifth-related dominant-seventh chords. (E) A voice leading between tritone-related dominant-seventh chords, related to (D) by individual T. (F) A voice leading between tritone-related half-diminished sevenths, related to (D) by individual T and uniform I.
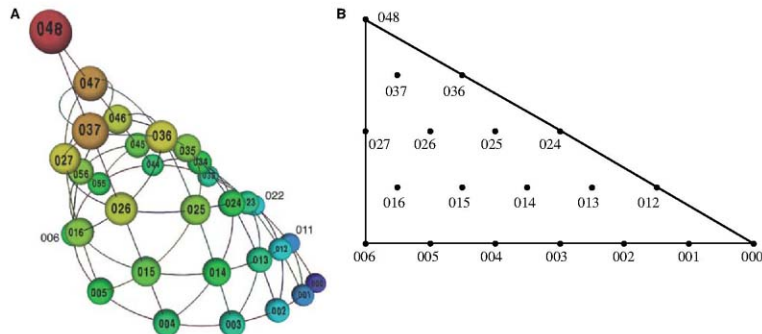

Fig. 2. (A) $\mathbb{T}^2/\mathcal{S}_3$ is a cone. (B) $\mathbb{T}^2/(\mathcal{S}_3 \times \mathbb{Z}_2)$ is a triangle. Numbers refer to pitch classes, with 0 = C, 1 = C#, etc. Points represent equivalence classes of transpositionally (A) or transpositionally and inversionally (B) related chords. Thus, (C, D, E) and (D, E, F#) are both instances of 024.

reflection. These spaces can be visualized as oblique cones over quotients of the $(n − 2)$-sphere, albeit with additional orbifold points. The singular point at the cone's "vertex" contains the chord type that divides the octave into n equal pieces; the more closely a chord type's notes cluster together, the farther it is from this point. The singular base of conical OPT space can be visualized as the quotient of the $(n − 2)$-sphere by the cyclic group $\mathbb{Z}_n$. (When n is prime, this is a lens space; when n is not prime, it does not in general seem to have a familiar name.) The base acts like a mirror, containing chords with note duplications. The OPTI-spaces $\mathbb{T}^{n−1}/(S_n \times \mathbb{Z}_2)$ are essentially similar: They can be visualized as cones over the quotients of the $(n − 2)$-sphere by the dihedral group $\mathbb{Z}_n \rtimes \mathbb{Z}_2$, with $\mathbb{Z}_2$ representing central inversion.

Figure 2A depicts $\mathbb{T}^2/\mathcal{S}_3$, the space of three-note chord types (OPT equivalence classes). The augmented triad at the vertex, 048, divides the octave perfectly evenly; major and minor triads, 047 and 037, are found near the tip and, as the basic sonorities of Western tonality, the unison 000 occupies the "kink" in the cone's singular base, which acts like a mirror. Orthogonal projection creates a barycentric coordinate system, seen here as a triangular grid (1). Pairs of points represent successions of chord types, whereas line segments represent OPT voice-leading classes (Fig. 1). Figure 2B is $\mathbb{T}^2/(\mathcal{S}_3 \times \mathbb{Z}_2)$, the quotient of the cone by a reflection. Points are OPTI equivalence classes (set classes), line segments are OPTI voice-leading classes, and all three boundaries act like mirrors. Figure S5 depicts the analogous four-note structures. Because the spaces are conical, some line segments near the vertex will self-intersect; musically, this means that near-

ly even chords can be linked to their transpositions by efficient voice leading (4).

The advantage of these constructions is that they permit a continuous generalization of traditional music-theoretical terminology. Informal musical discourse recognizes degrees of relatedness: Equal-tempered and just-intonation major triads are considered highly similar, even though they are not related by any OPTIC transformations. Likewise, composers often use scales in which scalar transposition [translation along a scale (4)] is nearly equal to log-frequency transposition: In such scales, fragments such as C-D-E ("Do, a deer") and D-E-F ("Re, a drop") are considered similar, even though they are not OPTIC-equivalent. Traditional music theory, however, has often adopted a binary approach to classification: Chords are considered equivalent if they can be related by OPTIC transformations and are considered unrelated otherwise (2). Several theorists have recently criticized this view, and modeling similarity between chord types is an active area of music-theoretical research (4, 11). However, no existing model describes the broad flexibility inherent in ordinary musical terminology; for example, none explains the similarity between just and equal-tempered major triads.

Our spaces suggest such a measure. Nearby points represent equivalence classes whose members can be linked by small voice leadings; in this sense, they are nearly equivalent modulo the relevant equivalence relation. This model is in good accord with traditional musical practice. Figure 3 shows that the traditional musical term "scale fragment" refers to a small region of the graph of equal-tempered chords in three-note conical OPT space ($\mathbb{T}^2/\mathcal{S}_3$); likewise, the term "triad" refers to a region of the equal-tempered graph surrounding the

vertex. Because the various tunings of the major triad lie very close together in $\mathbb{T}^2/\mathcal{S}_3$, the term "major triad" can be taken to refer to an even smaller region of the continuous space.

Geometrical representations of musical relatedness can also be useful in specific analytical contexts. Figure 3B shows three fragments from an early Schoenberg piano piece, which traditional theory would consider to be unrelated. Some theorists (12) have noted that the fragments can be understood as variations on the same musical idea. Following Straus (13), we can model Schoenberg's variations geometrically: Fig. 3A shows that the fragments define a sequence of short moves in $\mathbb{T}^2/\mathcal{S}_3$, each producing small changes in the motive's basic intervals. Similar analytical techniques can be used to compare progressions, rather than isolated chords (4). Because the OPTIC spaces are continuous, these techniques can potentially be applied to non-Western and microtonal music as well.

Beyond modeling music similarity, the geometrical perspective provides a unified framework for investigating a wide range of contemporary music-theoretical topics, including "contour" and "K-nets" (4, 14, 15). This reflects the fact that the OPTIC equivalences have been central to Western musical discourse since at least the seventeenth century (16). Our model translates these music-theoretical terms into precise geometrical language, revealing a rich set of mathematical consequences. This translation may have implications for theory, analysis, pedagogy, composition, musical data analysis and visualization, and perhaps even the design of new musical instruments.
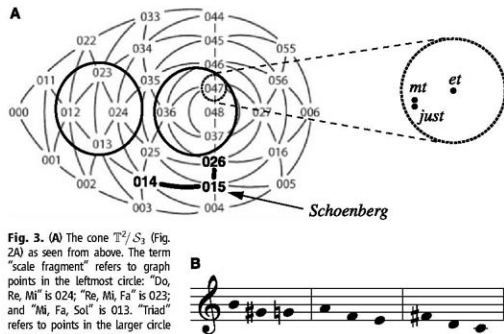
References and Notes

1. C. Callender, *Music Theory Online* **10** (2004); available online at http://mto.societymusictheory.org/issues/mto.04.10.3/mto.04.10.3.callender.pdf.
2. D. Lewin, *Generalized Musical Intervals and Transformations* (Yale, New Haven, CT, 1987).
3. G. Mazzola, *The Topos of Music* (Birkhäuser, Boston, MA, 2002).
4. Materials and methods are available as supporting material on *Science* Online.
5. D. Tymoczko, paper presented at the annual meeting of the Society for Music Theory, 12 November 2004, Seattle, WA.
6. R. Bott, *Fundam. Math.* **39**, 264 (1952).
7. C. Tuffley, *Alg. Geom. Topogr.* **2**, 1119 (2002).
8. Z. Ran, *Int. Math Res. Notices* **1993**, 93 (1993).
9. D. Tymoczko, *Science* **313**, 72 (2006).
10. R. Cohn, *Music Theory Online* **9** (2003); available online at http://mto.societymusictheory.org/issues/mto.03.9.4/mto.03.9.4.cohn.pdf.
11. I. Quinn, *Perspect. New Music* **39**, 108 (2001).
12. E. Haimo, *Music Theory Spectrum* **18** (1996).
13. J. Straus, *Music Theory Spectrum* **25**, 305 (2003).
14. R. Morris, *Music Theory Spectrum* **15**, 205 (1993).
15. H. Klumpenhouwer, thesis, Harvard University (1991).
16. J. P. Rameau, *Treatise on Harmony* (Dover, New York, 1971).
17. We thank N. Elkies, J. Ellenberg, D. Galtsgory, and L. Ramshaw.

Fig. 3. (A) The cone $\mathbb{T}^2/\mathcal{S}_3$ (Fig. 2A) as seen from above. The term "scale fragment" refers to graph points in the leftmost circle: "Do, Re, Mi" is 024; "Re, Mi, Fa" is 023; and "Mi, Fa, Sol" is 013. "Triad" refers to points in the larger circle on the right, containing major (047), minor (037), augmented (048), and diminished (036) chords. Different tunings of the major triad are found in the dotted circle around 047. On the right, the equal-tempered, just, and quarter-comma meantone triads. **(B)** Closely related motives at the beginning of Schoenberg's opus 11, no. 1. The variations define a pair of short line segments in the cone, from 014 to 015 to 026.

Picosecond Coherent Optical Manipulation of a Single Electron Spin in a Quantum Dot

J. Berezovsky,* M. H. Mikkelsen,* N. G. Stoltz, L. A. Coldren, D. D. Awschalom†

Most schemes for quantum information processing require fast single-qubit operations. For spin-based qubits, this involves performing arbitrary coherent rotations of the spin state on time scales much faster than the spin coherence time. By applying off-resonant, picosecond-scale optical pulses, we demonstrated the coherent rotation of a single electron spin through arbitrary angles up to π radians. We directly observed this spin manipulation using time-resolved Kerr rotation spectroscopy and found that the results are well described by a model that includes the electron-nuclear spin interaction. Measurements of the spin rotation as a function of laser detuning and intensity confirmed that the optical Stark effect is the operative mechanism.

Using ultrafast optical pulses to coherently manipulate the spin state of an electron is a key ingredient in many proposals for solid-state quantum information processing (1–6). Although electrical control of a single spin has been achieved (7, 8), the nanosecond time scales required for such manipulation limit the number of operations that can be performed within the spin coherence time. In contrast, spin control via picosecond-scale optical pulses yields an improvement of several orders of magnitude in the manipulation time, as required for many proposed applications. Additionally, this all-optical technique provides simple pathways for coupling the quantum degrees of freedom of a single spin and a photon—a useful property for enabling quantum communication. We experimentally demonstrated such a scheme for a single electron spin in a quantum dot (QD).

The optical (or ac) Stark effect (OSE) was first studied in atomic systems in the 1970s (9–11) and was subsequently explored in bulk semiconductors and quantum wells (12–14). In recent years, the OSE has been used to observe ensemble spin manipulation in a quantum well (15) and to control orbital coherence in a QD (16). Additionally, other optical manipulation schemes have been explored (17–20), which, in contrast to this work, affect the average polarization of a spin ensemble instead of directly manipulating individual spin states. Using perturbation theory, it is found that an optical field with intensity I_{tip} detuned from an electronic transition by an energy Δ, induces a shift in the transition energy

$$\Delta E \approx \frac{D^2 I_{tip}}{\Delta\sqrt{\varepsilon/\mu}} \quad (1)$$

where D is the dipole moment of the transition, and ε and μ are the permittivity and permeability of the material (13). In the relevant energy levels

for the QD system considered here (Fig. 1B), the ground state consists of a single electron in the lowest conduction band level, spin-split by a small magnetic field B_z. The lowest-energy interband transition is to the trion state, consisting of two electrons in a singlet state and a heavy hole. Because of the optical selection rules, the dipole strength of this transition in the basis along the $\hat{y}$ direction from the spin-up ground state is zero for σ^+ polarized light and from the spin-down ground state is zero for σ^- polarized light, as indicated in the diagram. Therefore, for circularly polarized light, the OSE shifts just one of the spin sublevels and produces a spin splitting in the ground state that can be represented as an effective magnetic field, B_{Stark}, along the light-propagation direction. By using ultrafast laser pulses with high instantaneous intensity to provide the Stark shift, large splittings can be obtained to perform coherent manipulation of the

spin within the duration of the optical pulse (here, $B_{Stark} \sim 10$ T). This phenomenon can also be described in terms of a stimulated Raman transition (3, 6, 20).

The sample used consisted of a layer of charge-tunable gallium arsenide (GaAs) interface QDs embedded in an optical cavity. [Details of the sample structure and characterization are given in (21–23).] In the experimental setup (Fig. 1C), three synchronized, independently tunable optical pulse trains are focused onto the sample: the pump, the probe, and the tipping pulse (TP). The pump and tipping pulse are both derived from a mode-locked Ti:sapphire laser generating a train of ~150-fs pulses at a repetition rate of 76 MHz. The pump is circularly polarized and tuned to an energy $E = 1.646$ eV [full width at half maximum (FWHM) ~ 1 meV], thereby injecting spin-polarized electrons and holes into the continuum of states above the QD (24). One or more of these electrons or holes can then relax into the QD. The circularly polarized TP (duration ~ 30 ps, FWHM = 0.2 meV) is tuned to an energy below the lowest QD transition (Fig. 1A) and is used to induce the Stark shift. The relative time delay between the pump pulse and the TP is controlled by a mechanical delay line in the pump path.

The probe pulse is generated by passing a narrow-linewidth continuous-wave laser through an electro-optic modulator synchronized with the pump/TP laser. The resulting 1.5-ns pulses probe the spin in the QD through the magneto-optical Kerr effect (23). This effect arises from the real part of the dielectric response function of the QD and results in a spin-dependent rotation of the polarization of the linearly polarized probe upon reflection off of the sample. As the energy of this probe light is scanned across the QD transi-

Fig. 1. (A) The solid red line shows trion (X⁻) and biexciton (XX) photoluminescence from a single QD. The black data points show the corresponding single-spin KR, and the black line is an odd-Lorentzian fit to this data, from which the KR amplitude θ_0 is obtained. **(B)** Diagram of the singly charged QD ground state and the optically allowed transitions to the trion state. The solid circles represent electrons and the open circles represent heavy holes. **(C)** Schematic of the experimental setup. CW, continuous wave; EOM, electro-optic modulator; pol. BS, polarizing beamsplitter.
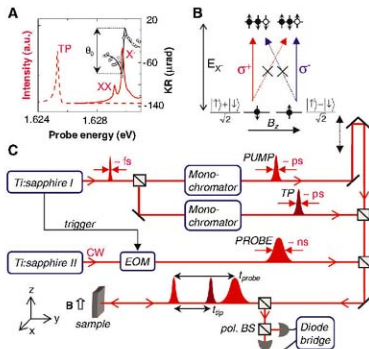
Center for Spintronics and Quantum Computation, University of California, Santa Barbara, CA 93106, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: awsch@physics.ucsb.edu

tion energy, E_X, an odd-Lorentzian line shape $[\propto x/(1 + x^2)]$ centered at the transition energy is seen in the Kerr rotation (KR) spectrum (22, 25). By fitting such a curve (as shown in Fig. 1A), we can extract the amplitude of this feature, θ_0, which is proportional to the projection of the spin polarization in the QD along the light-propagation direction.

In a typical measurement, the pump pulse arrives at time (t) = 0 along the y axis (growth direction), and in some cases, a single spin-polarized electron will relax into the QD. For pump helicity $\sigma^{\pm}$, this electron is (up to a global phase) initially in the state $\psi(t = 0) = (|\uparrow\rangle \pm i|\downarrow\rangle)/\sqrt{2}$, where "up" and "down" are chosen as the basis along the external magnetic field B_z. The spin then begins to coherently precess at the Larmor frequency $\omega = g\mu_B B_z/\hbar$: $\psi(t) = (\exp(-i\omega t/2)|\uparrow\rangle + i \exp(i\omega t/2)|\downarrow\rangle)/\sqrt{2}$, where g is the effective electron g factor, μ_B is the Bohr magneton, and $\hbar$ is the reduced Planck constant. At $t = t_{tip}$, the TP arrives and generates an additional spin splitting along the y axis for the duration of the pulse. During this time, the spin precesses about the total effective field (which is typically dominated by B_{swait}) and then continues to precess about the static applied field. The probe then measures the resulting projection of the spin in the QD, $\langle S_y \rangle$ at $t = t_{probe}$. This sequence is repeated at the repetition frequency of the laser (76 MHz), and the signal is averaged for several seconds for noise reduction. The pump and probe are modulated with mechanical choppers, allowing for lock-in detection to measure only spins that are injected by the pump. Furthermore, the pump is switched between σ^+ and σ^-, with a measurement made at each helicity. The spin signal is then taken as the difference between these values, eliminating any spurious signal from spins not generated by the pump [such as phonon-assisted absorption from the TP (26)].

To map out the coherent dynamics of the spin in the QD, KR spectra are measured as a function of pump-probe delay. Figure 2A shows a plot of the KR angle, θ_K as a function of probe energy and t_{probe} with an applied field $B_z = 715$ G and no TP. Horizontal line cuts display the dispersive line shape centered at the transition energy E_X (as shown in Fig. 1A). As t_{probe} is swept along the vertical axis, the precession of the spin can be seen as the oscillations in θ_K. When the TP is applied at $t_{tip} = 1.3$ ns, as in Fig. 2B, there is a substantial change in the KR spectra. For $t < t_{tip}$, the KR signal is essentially the same as in Fig. 2A, but for $t > t_{tip}$ the sign of the signal is reversed. This can be clearly seen in the line cuts in Fig. 2, C to F. Line cuts in (C) and (D) are both before the TP and show the same behavior, whereas the line cut in (F) has the opposite sign of the line cut in (E) as a result of the TP.

It is convenient to understand the observed spin dynamics in the Bloch sphere picture. Here, the spin state is represented as a vector (S_x, S_y, S_z), where (0, 0, $\pm$ S_z) represents the eigenstates $|\uparrow\rangle$ and $|\downarrow\rangle$, and vectors with nonzero S_x and S_y

represent coherent superpositions of $|\uparrow\rangle$ and $|\downarrow\rangle$. In this picture, the dynamics of the spin can be calculated by applying the appropriate sequence of rotation matrices to the initial state (21). The effect of the pump and the TP thus results in the spin state as a function of time

$$\vec{S}(t) =$$

$$\begin{cases} \begin{pmatrix} S_{0,y}\sin\omega t \\ S_{0,y}\cos\omega t \\ S_{0,z} \end{pmatrix} & 0 < t < t_{tip} \\[2em] \begin{pmatrix} S_{0,y}(\cos\phi_{tip}\sin\omega t_{tip}\cos\omega t' + \\ \cos\omega t_{tip}\sin\omega t') - \\ S_{0,z}\sin\phi_{tip}\cos\omega t' \\ S_{0,y}(-\cos\phi_{tip}\sin\omega t_{tip}\sin\omega t' + \\ \cos\omega t_{tip}\cos\omega t') - \\ S_{0,z}\sin\phi_{tip}\sin\omega t' \\ S_{0,y}\sin\phi_{tip}\sin\omega t_{tip} + S_{0,z}\cos\phi_{tip} \end{pmatrix} & t \geq t_{tip} \end{cases}$$

(2)

where the initial spin state is (0, $S_{0,y}$, $S_{0,z}$), $t' = t - t_{tip}$, ω is the precession frequency about the z axis, and ϕ_{tip} is the angle through which the spin precesses during the TP. Here we have included an initial component of the spin along the z axis to account for small misalignments of the pump beam. Because the duration of the TP is much less than ω^{-1}, the tipping is assumed to occur instantaneously.

As can be seen from Eq. 2, the TP may result in a substantial nonprecessing component of the spin along the z axis. Through the hyperfine interaction, electron spins can flip with nuclear spins. The component of the nuclear spin along the external field B_z does not precess and therefore can build over time in the process of dynamic nuclear polarization (DNP) (24, 27). As nuclear polarization builds along the z axis, it acts back on the electron spin as an effective magnetic field $B_n\hat{z}$. Thus in Eq. 2, ω must be replaced by $\omega' = \omega + \omega_n$. ω_n ($= g\mu_B B_n/\hbar$) is proportional to the steady-state nuclear polarization, which in turn is

proportional to S_z. Thus, equating ω_n and S_z with a constant of proportionality α, we have $\omega_n = \alpha(S_{0,y}\sin\phi_{tip}\sin(\omega + \omega_n)t_{tip} + S_{0,z}\cos\phi_{tip})$.

Solving this equation numerically for ω_n and substituting ω' into Eq. 2 self-consistently yield the coherent spin dynamics of the system as a function of time. To model the results below, we include the finite spin coherence time, T_2^*, and a phenomenological I_{tip}-dependent term to account for imperfect fidelity of the spin rotations or other background effects, with a characteristic scale I_0. The decoherence term is taken as exponential, in agreement with previous measurements on this sample (23). Finally, we take the difference between σ^+ and σ pump helicity, yielding

$$\theta_0(t; \vec{S}_0, \omega, I_{tip}, \alpha, T_2^*, I_0) = [S_y(t; \vec{S}_0) - S_y(t; -\vec{S}_0)]\exp(-t/T_2^*)\exp(-I_{tip}/I_0) \quad (3)$$

Figure 3A shows the time evolution of a single spin in a transverse magnetic field, with no TP applied. Each data point is determined from the fit to a KR spectrum at a given pump-probe delay, as in Fig. 2. If we convolve Eq. 3 with the measured profile of the probe pulse, we can perform a least-squares fit to this data and determine various parameters in the model: ω, T_2^*, and the effective field from nuclear polarization, B_n. The red curve in Fig. 3A shows the result of this fit, yielding $\omega = 1.39$ GHz, $T_2^* = 5.5$ ns, and $B_n = 68$ G, and the dotted line is the corresponding plot of Eq. 3 without the probe pulse convolution. As expected, the spin is initialized at $t = 0$ and then precesses freely about the applied field.

The data in Fig. 3, B and C, show the same coherent spin dynamics as those in Fig. 3A, but with the TP applied at $t = t_{tip}$. The intensity of the TP is chosen to induce a $1.05 \times \pi$ rotation about the y axis, which is determined as discussed below. In Fig. 3B, the TP arrives at $t_{tip} = 1.3$ ns, when the projection of the spin is mainly along the x axis. This component of the spin is thus rotated by the TP through $\sim\pi$ radians. The

Fig. 2. (**A** and **B**) KR as a function of probe energy and pump-probe delay. White represents positive KR, black represents negative KR, and the dashed blue line indicates the transition energy E_X. (A) shows spin precession in the absence of the TP; in (B), the TP is applied at $t_{tip} = 1.3$ ns, with $I_{tip} = 4.7 \times 10^5$ W/cm^2 and $\Delta = 2.65$ meV. A constant offset is subtracted from each KR spectrum. (**C** to **F**) Line cuts (black curves with data points) taken along the dashed red lines in (A) and (B) showing the effect of the TP. (C) and (E) are with no TP and $t_{probe} = 0$ and 2 ns, respectively; in (D) and (F), $t_{tip} = 1.3$ ns and $t_{probe} = 0$ and 2 ns, respectively. The solid red curves are odd-Lorentzian fits from which the KR amplitude θ_0 is extracted.
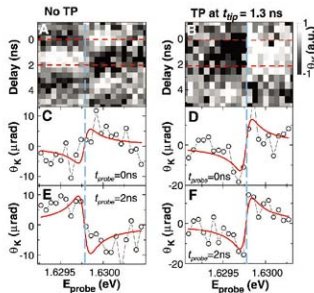
predicted spin dynamics as given by Eq. 3 is shown in the dotted red line, and the same curve convolved with the probe pulse is given by the solid red line. This curve is not a fit: All of the parameters are determined either in the fit to Fig. 3A or as discussed below. Only the overall amplitude of the curve has been normalized. Here, the spin is initialized at $t = 0$, and as be-

fore, precesses freely until the arrival of the TP. After the TP, the spin has been flipped and the resulting coherent spin dynamics show a reversal in sign. This can be clearly seen by comparing the sign of the measured signal at the position indicated by the dashed line in Fig. 3.

Figure 3C shows the spin dynamics again with the same parameters, but with $t_{tip} = 2.6$ ns. The spin at this delay will have only a small projection in the x-z plane, and the TP-induced rotation about the y axis will have only a small effect on the spin state. This expectation is borne out in the data, where the spin dynamics show essentially the same behavior as in the absence of the TP (Fig. 3A). Again, the model yields qualitatively the same behavior.

Further details of this spin manipulation can be investigated by varying the TP intensity I_{tip} and the detuning Δ of the TP from the QD transition energy for a fixed t_{probe} and t_{tip} (Fig. 4B). In Fig. 4A, the KR signal θ_0 as a function of I_{tip} is shown at $t_{probe} = 2.5$ ns, with the TP arriving at $t_{tip} = 1.3$ ns, for three different values of Δ. When $I_{tip} = 0$, the spin precesses undisturbed and yields a negative signal at $t_{probe} = 2.5$ ns (as in Fig. 3A). As I_{tip} is increased, the spin is coherently rotated through an increasingly large angle, and the observed signal at $t_{probe} = 2.5$ ns changes sign and becomes positive (as in Fig. 3B). Furthermore, the strength of the OSE is expected to decrease linearly with the detuning Δ (Eq. 1). The gray lines in Fig. 4A are plots of Eq. 3 with parameters taken from the fit in Fig. 3A, and $\phi_{tip} = \beta / I_{tip}$. From these curves, we additionally obtain the phenomenological factor $I_0 = 6.9 \times 10^5$ W/cm². This same value is used in all of the model curves shown. From this, we can estimate the fidelity of a π rotation to be approximately 80% (21). The only parameter that is changed between the three curves in Fig. 4A is the strength of the OSE, β. The TP intensity required for a π rotation can be extracted from these fits as $I_\pi = \pi/\beta$, and is shown in Fig. 4C as a function of detuning, displaying the expected linear dependence. If the expected slope of this line is estimated from Eq. 1 with reasonable parameters, we find a discrepancy of about a factor of 10 from the measured value, which may be due to a number of experimental uncertainties (21).

The data in Fig. 4A most clearly show the effects of DNP on the observed spin dynamics. In the absence of nuclear polarization, one would expect the curves in Fig. 4A to be cosinusoidal, crossing zero at an intensity half that required for a π rotation. DNP, however, which is maximal when $\phi_{tip} \approx \pi/2$, distorts this ideal cosine form, as is well described by the model (21). Additionally, an effect of the misalignment of the initial spin direction out of the x-y plane can be seen in the signal near $I_{tip} = 0$. The TP first rotates the spin into the x-y plane, increasing the signal slightly, and then rotates it past the x-y plane as I_{tip} is increased. This is confirmed by reversing the helicity of the TP so that the spin rotation is in the same direction as the misalignment, instead of
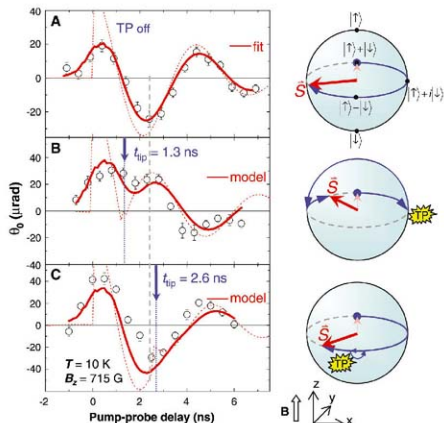

Fig. 3. (A) Coherent single-spin precession in a transverse magnetic field ($B_z = 715$ G), with error bars determined as the standard error from the fits to the KR spectra. The solid red line is a fit to Eq. 3 convolved with the probe pulse, and the dotted red line line shows Eq. 3 without the probe pulse convolution. The diagrams on the right schematically show the evolution of the spin on the Bloch sphere. (B and C) Spin dynamics under the same conditions as in (A), but with the TP applied at $t_{tip} = 1.3$ ns (B) and 2.6 ns (O) to induce a $1.05 \times \pi$ rotation about $\hat{y}$, with the TP detuning $\Delta = 2.65$ meV and intensity $I_{tip} = 4.7 \times 10^5$ W/cm². The solid red lines show model curves of Eq. 3 using parameters obtained elsewhere, with probe pulse convolution; the dotted red lines show model curves without probe pulse convolution.
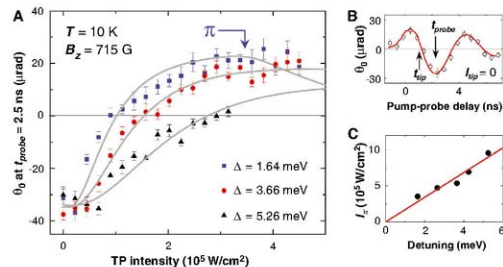

Fig. 4. (A) Single-spin KR amplitude θ_0 as a function of TP intensity, at three detunings from the X⁻ transition. The TP arrives at $t_{tip} = 1.3$ ns, and the probe is fixed at $t_{probe} = 2.5$ ns, as illustrated in (B). The gray lines are fits to Eq. 3, varying only a single parameter, the strength of the OSE (β), and using the parameters from the fit to Fig. 2A, $S_{0,z}/S_{0,y} = -0.11$ and $I_0 = 6.9 \times 10^5$ W/cm². The TP intensity, I_π, required for a π rotation at $\Delta = 1.64$ meV is indicated by the arrow. (C) I_π as a function of Δ, as obtained from the fits.

against it. [This and further control measurements are discussed in (21).]

Much of the deviation of the data from the model can be explained by the slow drift of experimental parameters during the measurement. In particular, the observed effects are very sensitive to the focus on the sample, because the intensities of the pump, probe, and TP all vary quadratically with the focused spot size. Additional deviations may be due to the simplistic description of the TP-induced background effects used here. For example, in the case of phonon-assisted transitions to the trion state, one would expect the type of spin-selective decoherence described in (18). Although there is some finite probability for the TP to excite the trion state, the control measurements described in the supporting online text show that TP-induced spin coherence is not the dominant mechanism for the spin control observed here. Further measurements of the background effects will be needed to determine their cause, with the aim of increasing the fidelity of these single spin rotations.

In principle, at most 200 single-qubit flips could be performed within the measured T_2^* of 6 ns. However, by using shorter TPs and QDs with longer spin coherence times, this technique could be extended to perform many more operations within the coherence time. A mode-locked laser producing ~100-fs TPs could potentially exceed the threshold (~10^4 operations) needed for proposed quantum error-correction schemes (28). Additionally, the spin manipulation demonstrated here may be used to obtain a spin echo (29), possibly extending the observed spin coherence time. These results represent progress toward the implementation of scalable quantum information processing in the solid state.

References and Notes

1. A. Imamoglu et al., Phys. Rev. Lett. 83, 4204 (1999).
2. M. Combescot, O. Betbeder-Matibet, Solid State Commun. 132, 129 (2004).
3. P. Chen, C. Piermarocchi, L. J. Sham, D. Gammon, D. G. Steel, Phys. Rev. B 69, 075320 (2004).
4. C. E. Pryor, M. E. Flatté, Appl. Phys. Lett. 88, 233108 (2006).
5. S. E. Economou, L. J. Sham, Y. Wu, D. G. Steel, Phys. Rev. B 74, 205415 (2006).
6. S. M. Clark, K.-M. C. Fu, T. D. Ladd, Y. Yamamoto, Phys. Rev. Lett. 99, 040501 (2007).
7. F. H. L. Koppens et al., Nature 442, 766 (2006).
8. K. C. Nowack, F. H. L. Koppens, Y. V. Nazarov, L. M. K. Vandersypen, Science 318, 1430 (2007).
9. C. Cohen-Tannoudji, J. Dupont-Roc, Phys. Rev. A 5, 968 (1972).
10. C. Cohen-Tannoudji, S. Reynaud, J. Phys. B 10, 345 (1977).
11. D. Suter, H. Klepel, J. Mlynek, Phys. Rev. Lett. 67, 2001 (1991).
12. M. Combescot, R. Combescot, Phys. Rev. Lett. 61, 117 (1988).
13. M. Joffre, D. Hulin, A. Migus, M. Combescot, Phys. Rev. Lett. 62, 74 (1989).
14. G. Papageorgiou et al., Phys. Rev. B 69, 085311 (2004).
15. J. A. Gupta, R. Knobel, N. Samarth, D. D. Awschalom, Science 292, 2458 (2001).
16. T. Unold, K. Mueller, C. Lienau, T. Elsaesser, A. D. Wieck, Phys. Rev. Lett. 92, 157401 (2004).
17. M. V. Gurudev Dutt et al., Phys. Rev. B 74, 125306 (2006).
18. Y. Wu et al., Phys. Rev. Lett. 99, 097402 (2007).
19. A. Greilich et al., Phys. Rev. B 75, 233301 (2007).
20. S. G. Carter, Z. Chen, S. T. Cundiff, Phys. Rev. B 76, 201308 (2007).
21. Supporting material is available on Science Online.
22. J. Berezovsky et al., Science 314, 1916 (2006).
23. M. H. Mikkelsen, J. Berezovsky, N. G. Stoltz, L. A. Coldren, D. D. Awschalom, Nat. Phys. 3, 770 (2007).
24. F. Meier, B. P. Zakharchenya, Eds., Optical Orientation: Modern Problems in Condensed Matter (North Holland, Amsterdam, 1984).
25. A. Atatüre, J. Dreiser, A. Badolato, A. Imamoglu, Nat. Phys. 3, 101 (2007).
26. A. Von Lehmen, J. E. Zucker, J. P. Heritage, D. S. Chemla, Phys. Rev. B 35, 6479(R) (1987).
27. D. Gammon et al., Phys. Rev. Lett. 86, 5176 (2001).
28. D. D. Awschalom, D. Loss, N. Samarth, Eds., Semiconductor Spintronics and Quantum Computation (Springer-Verlag, Berlin, 2002).
29. M. Rosatzin, D. Suter, J. Mlynek, Phys. Rev. A 42, 1839(R) (1990).
30. We acknowledge support from NSF and the Air Force Office of Scientific Research.

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5874/349/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
References

15 October 2007; accepted 26 February 2008
10.1126/science.1154798

Coherent Dynamics of a Single Spin Interacting with an Adjustable Spin Bath

R. Hanson,[1]*† V. V. Dobrovitski,[2] A. E. Feiguin,[1] O. Gywat,[1] D. D. Awschalom[1]

Phase coherence is a fundamental concept in quantum mechanics. Understanding the loss of coherence is paramount for future quantum information processing. We studied the coherent dynamics of a single central spin (a nitrogen-vacancy center) coupled to a bath of spins (nitrogen impurities) in diamond. Our experiments show that both the internal interactions of the bath and the coupling between the central spin and the bath can be tuned in situ, allowing access to regimes with surprisingly different behavior. The observed dynamics are well explained by analytics and numerical simulations, leading to valuable insight into the loss of coherence in spin systems. These measurements demonstrate that spins in diamond provide an excellent test bed for models and protocols in quantum information.

Quantum systems interact with their environment, resulting in a loss of initial coherence over time (1). Such system-bath interactions are studied extensively in a few canonical examples such as the spin-boson model (2) and the central spin model. In the latter, the coherence of a single spin (the central spin) in contact with a bath of spins is investigated (3–11). Study of the central spin problem may shed light on the emergence of the classical world from a collection of interacting quantum systems (1). Moreover, understanding spin-bath interactions is crucial for using spins in solids for quantum information processing (12–14), in which the efficient isolation of single quantum systems from their environment is required.

Studies in the field of nuclear magnetic resonance (NMR) and electron spin resonance have yielded detailed information about magnetic interactions in ensembles of spins (15). Recently, it has become possible to detect and coherently control individual spins (16, 17), allowing studies of the central spin model on truly single spins and possible applications in high-resolution magnetometry (18). We report here on a detailed study of the coherent dynamics of a single spin of a nitrogen-vacancy (NV) center in contact with a bath of nitrogen (N) impurity spins in diamond.

NV centers are well suited for studying spin interactions: Their spin state can be optically imaged, initialized, and read out, as well as controlled with high fidelity. In ultrapure diamond, the spin coherence time reaches hundreds of microseconds, being limited only by the weak interactions with nuclear spins of carbon-13 (19, 20). Therefore, the presence of nearby electron spins in diamond, even if few in number, can strongly influence the NV center spin, as the magnetic moment of an electron spin is three orders of magnitude larger than that of a nuclear spin.

In type Ib diamonds, as studied here, the magnetic environment of an NV center is dominated by N impurities (21), which carry an electronic spin of 1/2. These N spins are not optically active themselves but can be detected through the magnetic dipolar coupling with the NV center spin (22, 23). Previously, spin pairs were studied in which the dynamics of a single NV center spin were dominated by a single nearby N spin (19, 24). We studied the opposite regime, where the central spin (the NV center) is

[1]California Nanosystems Institute, University of California, Santa Barbara, CA 93106, USA. [2]Ames Laboratory and Iowa State University, Ames, IA 50011, USA.

*Present address: Kavli Institute of Nanoscience Delft, Delft University of Technology, Post Office Box 5046, 2600 GA Delft, Netherlands.
†To whom correspondence should be addressed. E-mail: r.hanson@tudelft.nl

interacting with a bath of N spins. Although the spin bath extends over the whole diamond, the dynamics of each individual NV spin are mainly determined by its local environment of N spins. Therefore, the ability to image and manipulate single NV centers (*16*) is crucial for these studies, because variations within an ensemble can average out many of the interesting dynamics.

A NV center consists of a substitutional N atom with an adjacent vacant site (V) in the diamond lattice (Fig. 1A). Its electronic ground state is a spin triplet ($S = 1$), with an energy splitting D of 2.87 GHz between states $m_S = 0$ and $m_S = \pm 1$ due to the crystal field (m_S is the projection of the spin on the z axis) (Fig. 1B). We imaged single NV centers at room temperature using a confocal microscope (Fig. 1C). The NV spin is first optically pumped into the $m_S = 0$ sublevel (Fig. 1D). Then, pulsed radiofrequency radiation is used to coherently manipulate the spin in the dark. Finally, readout is performed by measuring the photoluminescence rate, which

reflects the spin state (*16*, *23*). This cycle is typically repeated 10^5 times to build up statistics. The photoluminescence signal is normalized using the signal levels right after the initialization [when $p(m_S = 0) = 1$, where $p(m_S = 0)$ is the probability to be in the state $m_S = 0$] and after a π pulse [when $p(m_S = 0) = 0$].

We first showed that the spin bath properties can be controlled by subjecting the diamond to different static magnetic fields B. The total static field h_z acting on the NV center spin is the sum of B and the crystal-field splitting D (which can be viewed as an effective magnetic field) (*25*). We characterized the spin bath using standard NMR pulse sequences (*15*). We first measured the dephasing of the NV center spin during free evolution (Fig. 2A). We observe precession of the electron spin due to the hyperfine interaction with the N nuclear spin $I = 1$ of the NV center. This hyperfine shift is essentially static because of the large nuclear quadrupole splitting and slow nuclear spin relaxation. The N spins near

the NV center create an additional field through the magnetic dipolar interactions. This bath field δh drifts with time because of N spin flips, so that the NV center experiences a different field every time a new pulse sequence is started. After averaging over many sequences, which is required to build up statistics, this drift causes rapid decay of the free evolution signal, even though δh may fluctuate only very slowly on the time scale of a single pulse sequence (that is, quasi-static dephasing). The component of δh directed along the static field h_z, δh_z, has a much larger effect than the components perpendicular to h_z (such as δh_x and δh_y). This can be seen by transforming to the rotating frame of the spin (Fig. 2B). Here δh_x and δh_y are averaged out by fast rotations around h_z, whereas δh_z is unaffected by the transformation. The damping of the free evolution has a Gaussian shape, indicating that the distribution of δh_z, $P(\delta h_z)$, is also Gaussian: $P(\delta h_z) = 1/\sqrt{2\pi b^2}\exp(-\delta h_z^2/2b^2)$ (*15*). Its standard deviation b can be extracted from the decay during free evolution (*25*). Values for b are in the range of 0.3 to 1.1 MHz for four NV centers investigated, which is in good agreement with the average separation between N spins in this diamond of a few nanometers.

The static dephasing can be canceled with a spin echo (Fig. 2C). The time scale T_2 on which the spin echo signal decays (the spin coherence time) is proportional to the fluctuation rate of the spin bath (*15*). T_2 is almost an order of magnitude longer at $B = 740$ G than at $B = 0$ G, revealing a drastic change in the bath dynamics upon application of a magnetic field. This is explained by the dependence of the energy levels on the magnetic field (Fig. 2D). Close to $B = 0$ G, the average dipolar coupling between the N spins is larger than the energy splitting between the spin states, causing rapid fluctuations in the spin orientations (Fig. 2E). An applied magnetic field

Fig. 1. Detection and manipulation of a single spin. (**A**) Structure of an NV center in the diamond lattice. (**B**) The level structure of the NV center. The NV center is optically excited with a laser at 532 nm. Photoluminescence (PL) is measured with a confocal microscope. (**C**) Spatial PL map of part of the diamond sample, showing bright spots corresponding to single NV centers. (**D**) Measurement cycle consisting of initialization for 3 μs, manipulation through rf radiation that rotates the spin with Rabi frequencies of 5 to 30 MHz, and readout for 2 μs.
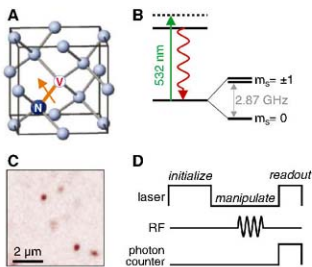

Fig. 2. Characterization of the spin bath. (**A**) Measurement of dephasing during free evolution of NV14. The Ramsey pulse scheme used is π/2-τ-π/2, where τ is the free evolution time. At $B = 0$ G, a superposition of $m_S = 1$ and $m_S = -1$ is prepared, whereas at $B = 740$ G, a superposition of $m_S = -1$ and $m_S = 0$ is prepared, resulting in opposite values at τ = 0 (*24*). Fitting to a Gaussian envelope decay yields $b = (0.42 \pm 0.03)$ MHz at $B = 0$ G and $b = (1.09 \pm 0.15)$ MHz at $B = 740$ G. (**B**) Bloch sphere representation of dephasing due to fluctuations δh of the total magnetic field h_z. In the rotating frame of the spin, only fluctuations along the magnetic field; that is, δh_z are significant. (**C**) Measurement of T_2 of NV14 using a π/2-τ-π-τ-π/2 spin echo pulse scheme. Fitting to exp($-2\tau/T_2$) yields $T_2 = (0.39 \pm 0.04)$ μs at $B = 0$ G and $T_2 = (3.2 \pm 0.4)$ μs at $B = 740$ G. (**D**) Energy of a single N electron spin (top) and a single NV center electron spin (bottom) as a function of B. (**E**) Schematic of the spin-bath dynamics at $B = 0$ G and $B = 740$ G.
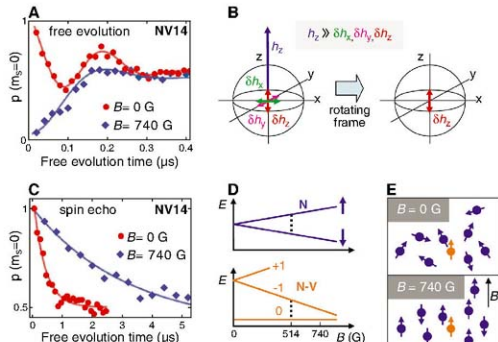
induces a large Zeeman energy splitting, which freezes out most of the spin dynamics (25). These experiments demonstrate that by tuning the magnetic field, we can control the dynamics within the spin bath.

We studied the spin-bath dynamics in more detail by measuring coherently driven spin oscillations (Rabi oscillations) at different magnetic fields. At $B = 0$ G (Fig. 3A), the oscillations initially decay fast and collapse almost completely, revive, and finally damp out slowly. This com-

plex behavior is reproduced at other NV centers and is observed for different driving fields.

To gain insight into these dynamics, we used analytical calculations and numerical simulations, based on existing knowledge about the internal structure of NV centers and N impurities (26). The dynamics of the NV spin were simulated numerically, using six N impurities at random locations, with the local density of N impurities being the only unknown parameter. This density was adjusted to match the data from the Ramsey

measurement (25). We explicitly take the nuclear spins of the nitrogen impurities into account, so that every impurity is, in fact, a system of two coupled spins (the electron with spin $S = 1/2$ and the nucleus with $I = 1$). Also, an analytical description was constructed by modeling the spin bath as a random field acting on the NV spin (25, 27, 28). The static and dynamical components of this field are characterized by the parameters b and T_2, respectively, whose values are known from the Ramsey and spin echo measure-

Fig. 3. Rabi oscillations at 0 G. (**A**) (Top) Rabi oscillations of NV14 at $B = 0$ G. (Middle) Analytical calculation of Rabi oscillations at $B = 0$ G for $b = 0.42$ MHz and a driving frequency of 16.6 MHz. (Bottom) Numerical simulation of Rabi oscillations at $B = 0$ G using a bath of six N spins. (**B**) (Left) Depiction of the magnetic fields in the rotating frame for the case of Rabi oscillations driven by an on-resonance field h_x. (Right top) Gaussian distribution of magnetic fields along z in the rotating frame with $b = 0.45$ MHz. (Right bottom) Rabi oscillations of a two-level system for $h_x = 8$ MHz, numerically averaged over the field distribution from the right top panel. (**C**) Same as (B), but with a static offset Δh_z of 2.3 MHz along z.
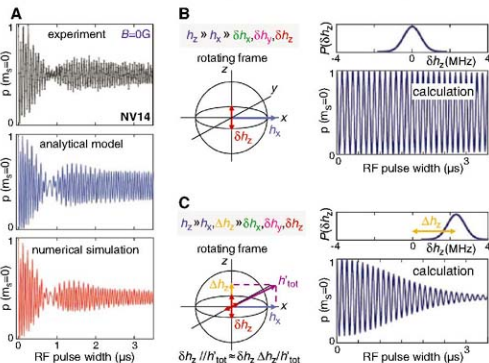
Fig. 4. Rabi oscillations at 740 and 514 G. (**A**) Rabi oscillations of NV14 at $B = 740$ G. (**B**) Distribution of fields along z in the rotating frame, for $b = 1.1$ MHz and $\Delta h_z = 0, \pm 2.3$ MHz. The thick blue line is the sum of the three distributions. (**C**) Rabi oscillations of a two-level system for $h_x = 10$ MHz, numerically averaged over the total field distribution in (B). (**D**) Rabi oscillations of NV14 and (**E**) NV31 at $B = 514$ G. Red lines are fits to an exponentially damped sum of two cosines with different frequencies. (**F**) Fast Fourier transforms (FFT) of curves in (D) (left) and (E) (right), revealing two dominant nutation frequencies. Ampl., amplitude; arb., arbitrary. (**G**) Plot of the lower nutation frequency versus the higher nutation frequency derived from Rabi oscillations at $B = 499$ G (triangles), $B = 514$ G (circles), and $B = 530$ G (squares). Different colors correspond to different NV centers. The gray line highlights the universal proportionality factor of $\sqrt{2}$.
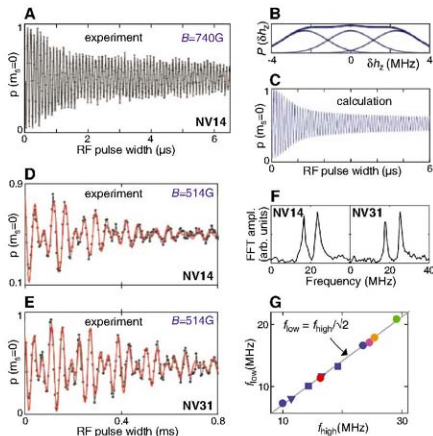
ments. For the Rabi oscillations, the dynamical component is neglected in the analytical model. Both the analytics and the numerical simulations accurately reproduce the essential features of the Rabi oscillations at $B = 0$ G without any fitting parameters (Fig. 3A): collapse and revival of the amplitude, beating of the oscillations at short times, and slow power-law decay after the revival.

From the theoretical analysis we find that the N nuclear spin of the NV center plays an essential role in the observed dynamics. Figure 3B depicts the fields acting on the spin in the frame rotating at frequency h_z, in case the nuclear spin is in the $m_I = 0$ state. Besides the fluctuating bath field δh_z, there is a strong driving field h_x, which rotates at the Larmor frequency set by h_z and thus appears static in the rotating frame. This case is equivalent to that shown in Fig. 2B, with h_z replaced by h_x and δh_z replaced by δh_x. Because δh_z is perpendicular to h_x, it is averaged out by fast spin precession around h_x and therefore has a very small effect on the dynamics: Rabi oscillations in this case show a slow power-law decay (4, 9, 10). In contrast, if the nuclear spin is in the $m_I = \pm 1$ state (Fig. 3C), a large offset Δh_z results because of the hyperfine interaction that tilts the sum of the fields, h'_{tot}, toward the z axis. In this case, the bath field δh_z has a component parallel to the total field h'_{tot}. This component is not affected by the fast precession around h'_{tot}. Again there is a correspondence with Fig. 2B, but now δh_z is replaced by $\delta h_z (\Delta h_z / h'_{tot})$. Although the distribution of the bath field is the same as in Fig. 3B, the Rabi oscillations decay rapidly with a Gaussian envelope because of the component parallel to h'_{tot}.

Experimentally, we measured the average over the nuclear spin configurations. The spin nutation (Rabi) frequency in the case of $m_I = \pm 1$ is slightly larger than in the case of $m_I = 0$, by a factor $\sqrt{(h_x^2 + \Delta h_z^2)}/h_x$. Therefore, beating will occur upon averaging because of interference between the oscillations corresponding to $m_I = 0$ and $m_I = \pm 1$, resulting in repeated collapse and revival. Because the oscillations corresponding to $m_I = \pm 1$ die out rapidly, only a single collapse and revival are visible. After the revival, the remaining signal is completely due to the $m_I = 0$ oscillations.

Even though the spin echo experiment indicates that the bath already changes on a time scale of 0.4 μs, our analytical model assuming a static bath shows excellent agreement with the data. The reason for this robustness is that under continuous driving, the spin is insensitive to fluctuations with a frequency below the Rabi frequency.

At $B = 740$ G, the collapse and revival are not observed in the Rabi oscillations (Fig. 4A). Because of alignment of the spins along B, the width of the bath field distribution is different than at $B = 0$ (25). For the measured value for b of 1.1 MHz, the distributions for the three possible NV nuclear spin states strongly overlap (Fig. 4B). As a consequence, the interference between the distributions leading to collapse is absent (Fig. 4C). Instead, the oscillations corresponding to the ex-

treme values of δh_z decay fast, and those corresponding to the more central values lead to a slow $1/\sqrt{t_{pw}}$ decay [t_{pw} is the radio frequency (rf) pulse width] (4, 9, 10).

The central spin and the bath spins can be brought into energy resonance by applying a magnetic field that exactly compensates for the NV center's crystal field splitting (near $B = 514$ G, Fig. 2D) (22, 23, 29). Here, the N electron spins can exchange their spins resonantly with the central spin through mutual flip-flops, providing an additional, efficient path for decoherence. Moreover, at this magnetic field the N spins are also resonant with the driving field and therefore will undergo driven rotations.

Figure 4, D to E, show Rabi oscillations at $B = 514$ G for two different NV centers. The oscillations clearly decay much faster than at $B = 740$ G. Furthermore, the data can be well fit to an exponential decay, as opposed to the power-law decay that was observed at $B = 740$ G. These observations suggest that in this regime the resonant spin flip-flop mechanism indeed dominates the decay of the Rabi oscillations.

A closer look at the data in Fig. 4, D to E, reveals a pronounced beating pattern. From a Fourier analysis (Fig. 4F) we find that there are two dominant oscillation frequencies. In Fig. 4G we plot the lower of these two frequencies, f_{low}, versus the higher frequency, f_{high}, for different NV centers and for different experimental conditions. The two frequencies differ exactly by a factor of $\sqrt{2}$, for all five NV centers investigated and for all driving frequencies.

Because of the strong hyperfine interaction of the N electron spin with its nuclear spin (26), the resonance condition also occurs at $B = 499$ G and $B = 530$ G (22, 23). We observe the same beating pattern at these fields. At $B = 522$ G, where the central spin and the bath are not resonant, only the higher of the two oscillation frequencies is present. This indicates that the lower oscillation frequency is induced by the resonance condition.

For a spin of $S = 1$ (as the NV center has), the Rabi frequency is larger than for a spin of $S = 1/2$ (as the N spins have) for the same driving field by $\sqrt{2}$ (15, 25). Therefore, the observed factor of $\sqrt{2}$ strongly suggests that the coherent rotation of N spins is the cause of the beating pattern. Rotation of the N spins will cause the dipolar field at the NV center to oscillate with f_{low}, which could in turn rotate the NV spin. However, the beating pattern is not reproduced by simulations or analytics if equilibrium conditions are assumed, suggesting that strongly nonequilibrium effects such as a polarization of the spin bath and initial entanglement between the central spin and the bath may play a crucial role. More work is needed to investigate this.

The use of more complicated pulse sequences will allow new studies of fundamental quantum-mechanical models and tests of protocols for future quantum information-processing tasks. Moreover, this system can be used as a test bed for similar systems, such as an electron spin cou-

pled to a bath of nuclear spins in a quantum dot (17), which lack the tunability and level of coherent control demonstrated here. Finally, the spin baths in our sample were formed randomly during crystal growth. Using the recently gained ability to position single spins in diamond by ion implanting (30), spin environments may soon be studied that are engineered from scratch.

References and Notes

1. W. H. Zurek, *Rev. Mod. Phys.* **75**, 715 (2003).
2. A. J. Leggett et al., *Rev. Mod. Phys.* **59**, 1 (1987).
3. N. V. Prokof'ev, P. C. E. Stamp, *Rep. Prog. Phys.* **63**, 669 (2000).
4. V. V. Dobrovitski, H. A. De Raedt, M. I. Katsnelson, B. N. Harmon, http://arxiv.org/abs/quant-ph/0112053 (2001).
5. A. V. Khaetskii, D. Loss, L. Glazman, *Phys. Rev. B* **88**, 186802 (2002).
6. I. A. Merkulov, A. L. Efros, M. Rosen, *Phys. Rev. B* **65**, 205309 (2002).
7. R. de Sousa, S. Das Sarma, *Phys. Rev. B* **68**, 115322 (2003).
8. V. V. Dobrovitski, H. A. De Raedt, M. I. Katsnelson, B. N. Harmon, *Phys. Rev. Lett.* **90**, 210401 (2003).
9. F. M. Cucchietti, J. P. Paz, W. H. Zurek, *Phys. Rev. A* **72**, 052113 (2005).
10. J. M. Taylor, M. D. Lukin, *Quant. Info. Process.* **5**, 503 (2006).
11. W. Yao, R.-B. Liu, L. J. Sham, *Phys. Rev. Lett.* **98**, 077602 (2007).
12. D. Loss, D. P. DiVincenzo, *Phys. Rev. A* **57**, 120 (1998).
13. B. E. Kane, *Nature* **393**, 133 (1998).
14. D. D. Awschalom, D. Loss, N. Samarth, Eds., *Semiconductor Spintronics and Quantum Computation* (Springer, Berlin, 2002).
15. C. P. Slichter, *Principles of Magnetic Resonance* (Springer-Verlag, New York, 1990).
16. F. Jelezko, T. Gaebel, I. Popa, A. Gruber, J. Wrachtrup, *Phys. Rev. Lett.* **92**, 076401 (2004).
17. R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, L. M. K. Vandersypen, *Rev. Mod. Phys.* **79**, 1217 (2007).
18. B. M. Chernobrod, G. P. Berman, *J. Appl. Phys.* **97**, 014903 (2005).
19. T. Gaebel et al., *Nat. Phys.* **2**, 408 (2006).
20. L. Childress et al., *Science* **314**, 281 (2006).
21. T. A. Kennedy, J. S. Colton, J. E. Butler, R. C. Linares, P. J. Doering, *Appl. Phys. Lett.* **83**, 4190 (2003).
22. E. van Oort, P. Stroomer, M. Glasbeek, *Phys. Rev. B* **42**, 8605 (1990).
23. R. J. Epstein, F. M. Mendoza, Y. K. Kato, D. D. Awschalom, *Nat. Phys.* **1**, 94 (2005).
24. R. Hanson, F. Mendoza, R. J. Epstein, D. D. Awschalom, *Phys. Rev. Lett.* **97**, 087601 (2006).
25. See supporting material on *Science* Online.
26. J. H. N. Loubser, J. A. van Wyk, *Rep. Prog. Phys.* **41**, 1201 (1978).
27. G. M. Zhidomirov, K. M. Salikhov, *Sov. Phys. J. Exp. Theor. Phys.* **29**, 1037 (1969).
28. J. R. Klauder, J. P. Anderson, *Phys. Rev.* **125**, 912 (1962).
29. R. Hanson, D. Gywat, D. D. Awschalom, *Phys. Rev. B* **74**, 161203(R) (2006).
30. J. R. Rabeau et al., *Appl. Phys. Lett.* **88**, 023113 (2006).
31. We acknowledge support from the Air Force Office of Scientific Research and the Dutch Organization for Fundamental Research on Matter (FOM). A.E.F is a member of Microsoft Station Q and acknowledges support from the Microsoft Corporation. Work at the Ames Laboratory is supported by the U.S. Department of Energy, Basic Energy Sciences, under contract DE-AC02-07CH11358.

Chaotic Dirac Billiard in Graphene Quantum Dots

L. A. Ponomarenko,[1] F. Schedin,[1] M. I. Katsnelson,[2] R. Yang,[1] E. W. Hill,[1] K. S. Novoselov,[1]* A. K. Geim[1]

The exceptional electronic properties of graphene, with its charge carriers mimicking relativistic quantum particles and its formidable potential in various applications, have ensured a rapid growth of interest in this new material. We report on electron transport in quantum dot devices carved entirely from graphene. At large sizes (>100 nanometers), they behave as conventional single-electron transistors, exhibiting periodic Coulomb blockade peaks. For quantum dots smaller than 100 nanometers, the peaks become strongly nonperiodic, indicating a major contribution of quantum confinement. Random peak spacing and its statistics are well described by the theory of chaotic neutrino billiards. Short constrictions of only a few nanometers in width remain conductive and reveal a confinement gap of up to 0.5 electron volt, demonstrating the possibility of molecular-scale electronics based on graphene.

One of the most discussed and tantalizing directions in research on graphene (1, 2) is its use as the base material for electronic circuitry that is envisaged to consist of nanometer-sized elements. Most attention has so far been focused on graphene nanoribbons (3–12). An alternative is quantum dot (QD) devices that, as described below, can be made entirely from graphene, including their central islands (CIs), quantum barriers, source and drain contacts, and side-gate electrodes.

Our experimental devices were microfabricated from graphene crystallites prepared by cleavage on top of an oxidized Si wafer (300 nm of SiO₂) (1). By using high-resolution electron-beam lithography, we defined a 30-nm-thick polymethylmethacrylate (PMMA) mask that protected chosen areas during oxygen plasma etching and allowed us to carve graphene into a desired geometry. The inset in Fig. 1A shows one of our working devices that generally consisted of the CI of diameter D, connected via two short constrictions to wide source and drain regions; the devices also had side-gate electrodes [we often placed them ~1 μm away from the CI as explained in (13)]. The Si wafer was used as a back gate. The constrictions were designed to have equal length and width of 20 nm (13), and we refer to them as quantum point contacts (QPCs). They provided quantum barriers to decouple the CI from the contacts (14, 15). If necessary, by using further etching (after the devices were tested), we could narrow QPCs by several nm, exploiting the gradual etching away of the PMMA mask not only from top but also sideways. This allowed us to tune the resistance of QDs to a value of several hundred kilohm, i.e., much larger than resistance quantum h/e^2 (e is the electron charge, and h is Planck's constant),

which is essential for single-electron transport. Graphene QDs with features as small as 10 nm could be fabricated reliably with this approach. For even smaller D, irregularities in PMMA [~5 nm (16)] became comparable in size with the designed features and, unavoidably on this scale, we could only estimate the device geometry. The measurements were carried out using the standard lock-in technique with dc bias at temperatures T from 0.3 to 300 K. We used both side and back gates; the latter allowed extensive changes in the population of QD levels, whereas the former was useful for accurate sweeps over small energy intervals (13). The response to the side-gate potential differed for different devices but could be related to back-gate voltage V_g through a numerical factor. For consistency, all the data are presented as a function of V_g.

We have found three basic operational regimes for QDs, depending on their D. Our large devices exhibit (nearly) periodic Coulomb blockade (CB) resonances that at low T are separated by regions of zero conductance G (Fig. 1A). As T increases, the peaks become broader and overlap, gradually transforming into CB oscillations (Fig. 1B). The oscillations become weaker as G increases with carrier concentration or T, and completely disappear for G larger than $\sim 0.5e^2/h$ because the barriers become too transparent to allow CB. For the data in Fig. 1B, we have identified more than 1000 oscillations. Their periodicity, $\Delta V_g \approx 16$ mV, yields the capacitance between the back gate and CI, $C_g = e/\Delta V_g \approx$ 10 aF, which is close to $C_g \approx 2\varepsilon_0(\varepsilon + 1)D \approx 20$ aF, as expected for a disk placed on top of SiO₂ (dielectric constant $\varepsilon \approx 4$) and at a distance $h \geq D = 250$ nm from the metallic Si gate (in this case, C_g is nearly the same as for an isolated disk) (17). The difference by a factor of 2 can be accounted for in terms of screening by the contact regions (17).

The overall shape of the conductance curve $G(V_g)$ in Fig. 1B resembles that of bulk graphene but is distorted by smooth (on the scale of ΔV_g) fluctuations that are typical for

mesoscopic devices and are due to quantum interference (1–4, 18–20). Smooth variations in the CB peak height (Fig. 1A) are attributed to interference-induced changes in the barriers' transparency, as shown by studying individual QPCs (13). Furthermore, we have measured the dependence of CB on applied bias V_b and, from the standard stability diagrams (Coulomb diamonds), found the charging energy E_c. The lower inset in Fig. 1B shows such diamonds for $D \approx 250$ nm, which yields $E_c \approx 3$ meV and the total capacitance $C = e^2/E_c \approx 50$ aF. The rather large E_c implies that the CB oscillations in Fig. 1B
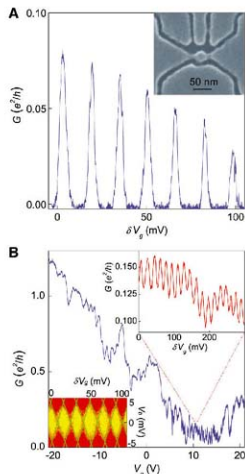

Fig. 1. Graphene-based single-electron transistor. **(A)** Conductance G of a device with the central island of 250 nm in diameter and distant side gates (13) as a function of V_g in the vicinity of +15 V (B); T = 0.3 K. The inset shows one of our smaller devices to illustrate the high resolution of our lithography that allows features down to 10 nm. Dark areas in the scanning electron micrograph are gaps in the PMMA mask so that graphene is removed from these areas by plasma etching. In this case, a 30-nm QD is connected to contact regions through narrow constrictions and there are four side gates. **(B)** Conductance of the same device as in (A) over a wide range of V_g (T = 4 K). Upper inset: Zooming into the low-G region reveals hundreds of CB oscillations. The lower inset shows Coulomb diamonds: differential conductance $G_{diff} = dI/dV$ as a function of V_g (around +10 V) and bias V_b (yellow-to-red scale corresponds to G_{diff} varying from zero to $0.3e^2/h$).

[1]Manchester Centre for Mesoscience and Nanotechnology, University of Manchester, Manchester M13 9PL, UK. [2]Institute for Molecules and Materials, Radboud University Nijmegen, 6525 ED Nijmegen, Netherlands.

*To whom correspondence should be addressed. E-mail: novoselov@manchester.ac.uk

are smeared mostly due to an increase in the barrier transparency with T, and submicrometer graphene QDs should be operational at $T \geq 10$ K. In general, the observed behavior is in agreement with the one exhibited by conventional single-electron transistors (SETs). Such devices were previously studied with metallic and semiconducting materials (14, 15) and, more recently, the first SETs made from graphite (18) and graphene (1, 19, 20) were also demonstrated. The all-graphene SETs reported here are technologically simple, reliable, and robust and can operate well above liquid-helium T, making them attractive candidates for use in various charge-detector schemes (14).

For devices smaller than ~100 nm, we observed a qualitative change in behavior: CB
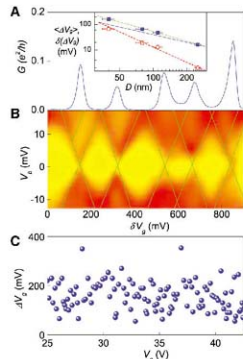
Fig. 2. Effect of quantum confinement. CB peaks (**A**) and Coulomb diamonds (**B**) for a 40-nm QD ($T = 4$ K). Variations in peak spacing and the size of diamonds are clearly seen. Yellow-to-red scale in (B) corresponds to G_{diff} varying from zero to $0.4e^2/h$. Two excited states (marked by additional lines) are feebly visible around $\delta V_g \approx 150$ and 850 mV and $V_b \approx 10$ mV. The smearing is caused by an increase in the transparency of quantum barriers at higher V_b. Further examples of excited states are given in the supporting material (13). (**C**) Separation of nearest-neighbor peaks at zero V_b in the same device for a large interval of V_g (beyond this interval, CB became suppressed by high transparency of the barriers). Inset in (A): Log-log plot of the average peak spacing $\langle \Delta V_g \rangle$ (solid squares) and its standard deviation $\delta(\Delta V_g)$ (open circles) as a function of D. Linear ($\langle \Delta V_g \rangle \propto 1/D$, solid line) and quadratic [$\delta (\Delta V_g) \propto 1/D^2$, dashed] dependences are plotted as guides to the eye. The dotted curve is the best fit for the average peak spacing: $\langle \Delta V_g \rangle \propto 1/D^\gamma$ yielding $\gamma = 1.25$.

peaks were no longer a periodic function of V_g but varied strongly in their spacing (Fig. 2, A and B), illustrate this behavior for $D = 40$ nm, whereas Fig. 2C plots the distance ΔV_g between the nearest peaks for 140 of them). One can see that ΔV_g varies by a factor of 5 or more, which exceeds by orders of magnitude typical variations of ΔV_g observed in nongraphene QDs (15, 21, 22). This is a clear indication that the size quantization becomes an important factor even for such a modest confinement. The reason for this is that a typical level spacing $\delta E \approx v_F h/2D$ for graphene's massless carriers [Dirac fermions (1, 2)] in a quantum box of size D is much larger than $\delta E \approx \hbar^2/8mD^2$ for massive carriers in other materials ($v_F \approx 10^6$ m/s is the Fermi velocity in graphene, and m is the effective mass). The distance between CB peaks is determined by the sum of charging and confinement energies $\Delta E = E_c + \delta E$, and the latter contribution becomes dominant for our devices with $D < 100$ nm. Accordingly, we refer to them as QDs rather than SETs (23). Because E_c is constant for a given QD geometry, variations in ΔE (found from the height of Coulomb diamonds) are due to confinement, which allows us to estimate characteristic δE (24). For example, we find $\delta E \sim 10$ meV for the 40-nm QD (Fig. 2), in agreement with the above formula $\delta E \approx \alpha/D$, where coefficient α varies around 0.5 eV·nm by a factor of 2 in different models (5–12).

For four devices with different D, we have carried out statistical analysis of their peak spacing (Fig. 3). As QDs become smaller, the average distance $\langle \Delta V_g \rangle$ between CB peaks gradually increases (Fig. 2A; inset). General expectations suggest that $\langle \Delta V_g \rangle$ should be proportional to $1/D$, being determined by two contributions to the QD capacitance: geometrical and quantum (15). According to the formula used above, the geometrical capacitance is $\propto D$. The quantum capacitance is given by the confinement energy and, in the first approximation, is also expected to vary as D. Indeed, it has been shown (25) that energy levels E_{nl} of Dirac fermions inside a disk of diameter D are described by $J_l(E_{nl}D/2\hbar v_F) = J_{l+1}(E_{nl}D/2\hbar v_F)$, where n and l are the main and

orbital quantum numbers, respectively, and $J_l(x)$ are the Bessel functions. This equation yields a typical level splitting ($\delta E) \propto 1/D$ (25), in qualitative agreement with the behavior of $\langle \Delta V_g \rangle$ in Fig. 2A.

However, further analysis reveals that the above simple picture starts to break down for $D < 100$ nm. One can see from Fig. 3 that the spectral distribution notably changes: For small QDs, the histograms become increasingly broader, as compared to their average positions. Also, $\langle \Delta V_g \rangle$ changes somewhat quicker than in $1/D$ (Fig. 2A, inset). We have calculated statistical deviations $\delta(\Delta V_g)$ from the average $\langle \Delta V_g \rangle$ and found that $\delta(\Delta V_g)$ grows approximately as $1/D^2$ with decreasing D (Fig. 2A). For example, for $D = 40$ nm, average fluctuations in the peak spacing become as large as $\langle \Delta V_g \rangle$ itself, which essentially means random positions of CB peaks. The observed behavior contradicts the one expected for Dirac fermions confined inside an ideal disk (25).

To explain this discrepancy, we point out that any confinement of Dirac fermions, except for the circular one, is predicted to result in quantum chaos (even the simplest square geometry leads to chaotic trajectories) (25). In general, chaos is a common feature of all systems with several degrees of freedom, whose behavior cannot be described as a superposition of independent one-dimensional motions. Classically, this entanglement between different variables leads to an exponential increase in the distance between two initially close trajectories with increasing time. Quantum mechanically, chaotic systems are characterized by distinctive statistics of their energy levels, which must comply with one of the Gaussian random ensembles, in contrast to the level statistics for the nonchaotic systems described by the Poisson distribution (26).

The experimentally observed level statistics in graphene QDs agrees well with the one predicted for chaotic Dirac or "neutrino" billiards (Fig. 3). Indeed, in our case, the quantum capacitance is no longer $\propto D$ because δE changes as $\propto 1/D^2$ (25), reflecting the lifting of the level
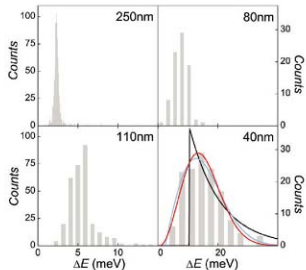

Fig. 3. Level statistics in Dirac billiards. Histograms of the nearest-neighbor level spacing in QDs of different D. The spacing is given directly in terms of δE (rather than ΔV_g, which was achieved by measuring the size of Coulomb diamonds. This allows the straightforward comparison between the experiment and theory (25). The level statistics becomes increasingly non-Poissonian for smaller QDs. This is illustrated for the smallest device where the red, black, and blue curves are the best fits for the Gaussian unitary, Poisson, and Gaussian orthogonal ensembles, respectively. There are no states at low ΔE because the measured distributions are shifted from the origin due to CB (E_c is used as a fitting parameter).

degeneracy at large n and l, and the number of states around a given energy is proportional to the dot area $\propto D^2$. This effect is often referred to as the level repulsion, a universal signature of quantum chaos. The observed random spacing of CB peaks, random height of Coulomb diamonds, changes in $\langle\Delta V_g\rangle$ quicker than $1/D$ and, especially, the pronounced broadening of the spectral distribution all indicate that chaos becomes a dominant factor for small QDs.

To corroborate this further, Fig. 3 shows that the observed level spacing is well described by Gaussian unitary distribution $(32/\pi^2)\delta E^2\exp(-48\delta E^2/\pi)$ (characteristic of chaotic billiards) rather than the Poisson statistics $\exp(-\delta E)$ expected for integrable geometries $(25, 26)$. The CB energy shifts the statistical distributions from zero (we measure $\Delta E = E_c + \delta E$ rather than δE), and this makes it difficult to distinguish between unitary and orthogonal ensembles. Nevertheless, the Gaussian unitary distribution fits our data notably better. This agrees with the theory that expects random edges to break down the sublattice symmetry (27) leading to the unitary statistics (25). The statistics of Dirac billiards are different from the chaotic wave systems that mimic quantum mechanics and are also described by the linear dis-

persion relation (optical, microwave, and acoustic cavities) but typically obey the Gaussian orthogonal statistics (28). Further evidence for the level repulsion in small QDs is provided by the absence of any apparent bunching in their spectra (Fig. 2C). Indeed, despite considerable effort, we did not find repetitive quartets or pairs of CB peaks, which in principle could be expected due to spin and/or valley degeneracy. The latter degeneracy is lifted by edge scattering (27), whereas the spin degeneracy may be removed by scattering on localized spins due to broken carbon bonds (5).

For even smaller devices $(D < 30nm)$, the experimental behavior is completely dominated by quantum confinement. They exhibit insulating regions in V_g sometimes as large as several V, and their stability diagrams yield the level spacing exceeding ~50 meV (Fig. 4, A and B). However, because even the state-of-the-art lithography does not allow one to control features <10 nm in size, the experimental behavior varies widely, from being characteristic of either an individual QD or two QDs in series or an individual QPC (13). It is also impossible to relate the observations with the exact geometry because scanning electron and atomic force microscopy fail in visualizing the one-atom-thick elements of several nm in size and often covered by PMMA or its residue. Nevertheless, we can still use δE to estimate the spatial scale involved. Basic arguments valid at a microscopic scale require $a/D \approx \delta E/t$ (where a is the interatomic distance, and $t \approx 3$ eV is the hopping energy), which again yields $\delta E \approx a/D$ with $a \approx 0.5$ eV nm. For example, for the QD shown in Fig. 4 with $\Delta E \approx 40$ meV, we find $D \sim 15$ nm.

Finally, we used our smallest devices (both QDs and QPCs) to increase δE by further decreasing their size using plasma etching. Some of the devices become overetched and stop conducting, but in other cases we have narrowed them down to a few nm so that they exhibit the transistor action even at room T (Fig. 4C). The device shown appears completely insulating, with no measurable conductance $(G < 10^{-10}$ S) over an extended range of V_g (>30 V) (off state), but then it suddenly switches on, exhibiting rather high $G \approx 10^{-2}e^2/h$. At large biases, we observe the conductance onset shifting with V_b (13), which allows an estimate for ΔE as ≈ 0.5 eV. This agrees with the T dependence measured near the onset of the on state, which shows that we do not deal with several QDs in series [as it was argued to be the case for nanoribbons (29)]. With no possibility to control the exact geometry for the nm sizes, we cannot be certain about the origin of the observed switching. Also, the exact boundary arrangements (armchair versus zigzag versus random edge and the termination of dangling bonds) can be important on this scale $(5-12)$. Nevertheless, $\delta E \sim 0.5$ eV again allows us to estimate the spatial scale involved in the confinement as only ~1 nm.

Our work demonstrates that graphene QDs are an interesting and versatile experimental

system allowing a range of operational regimes from conventional single-electron detectors to Dirac billiards, in which size effects are exceptionally strong and chaos develops easily. Unlike any other material, graphene remains mechanically and chemically stable and highly conductive at the scale of a few benzene rings, which makes it uniquely suitable for the top-down approach to molecular-scale electronics.

References and Notes

1. A. K. Geim, K. S. Novoselov, *Nat. Mater.* **6**, 183 (2007).
2. A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, *Rev. Mod. Phys.*, in press; preprint at http://xxx.lanl.gov/abs/0709.1163 (2007).
3. M. Y. Han, B. Ozyilmaz, Y. B. Zhang, P. Kim, *Phys. Rev. Lett.* **98**, 206805 (2007).
4. P. Avouris, Z. H. Chen, V. Perebeinos, *Nat. Nanotechnol.* **2**, 605 (2007).
5. Y. W. Son, M. L. Cohen, S. G. Louie, *Nature* **444**, 347 (2006).
6. D. Gunlycke, D. A. Areshkin, C. T. White, *Appl. Phys. Lett.* **90**, 142104 (2007).
7. L. Brey, H. A. Fertig, Y. W. Son, M. L. Cohen, S. G. Louie, *Phys. Rev. Lett.* **99**, 186801 (2007).
8. N. M. R. Peres, A. H. Castro Neto, F. Guinea, *Phys. Rev. B* **73**, 195411 (2006).
9. V. Barone, O. Hod, G. E. Scuseria, *Nano Lett.* **6**, 2748 (2006).
10. L. Brey, H. A. Fertig, *Phys. Rev. B* **73**, 235411 (2006).
11. B. Wunsch, T. Stauber, F. Guinea, *Phys. Rev. B* **77**, 035316 (2008).
12. J. Martin, Y. M. Blanter, preprint at http://lanl.arxiv.org/abs/0705.0532 (2007).
13. See supporting material on *Science* Online.
14. K. K. Likharev, *Proc. IEEE* **87**, 606 (1999).
15. L. P. Kouwenhoven *et al.*, in *Mesoscopic Electron Transport*, L. L. Sohn, L. P. Kouwenhoven, G. Schön, Eds. (Kluwer Series E345, Dordrecht, Netherlands, 1997), pp. 105–214.
16. E. A. Dobisz, S. L. Brandow, R. Bass, J. Mitterender, *J. Vac. Sci. Technol. B* **18**, 107 (2000).
17. D. Gelmont, M. S. Shur, R. J. Mattauch, *Solid State Electron.* **38**, 731 (1995).
18. J. S. Bunch, Y. Yaish, M. Brink, K. Bolotin, P. L. McEuen, *Nano Lett.* **5**, 287 (2005).
19. F. Miao *et al.*, *Science* **317**, 1530 (2007).
20. C. Stampfer *et al.*, *Appl. Phys. Lett.* **92**, 012102 (2008).
21. U. Sivan *et al.*, *Phys. Rev. Lett.* **77**, 1123 (1996).
22. S. R. Patel *et al.*, *Phys. Rev. Lett.* **80**, 4522 (1998).
23. P. G. Silvestrov, K. B. Efetov, *Phys. Rev. Lett.* **98**, 016802 (2007).
24. I. M. Ruzin, V. Chandrasekhar, E. I. Levin, L. I. Glazman, *Phys. Rev. B* **45**, 13469 (1992).
25. M. V. Berry, R. J. Mondragon, *Proc. R. Soc. London A* **412**, 53 (1987).
26. T. Gehr, A. Müller-Groeling, H. A. Weidenmüller, *Phys. Rep.* **299**, 189 (1998).
27. A. Rycerz, J. Tworzydło, C. W. J. Beenakker, *Nat. Phys.* **3**, 172 (2007).
28. U. Kuhl, H.-J. Stöckmann, R. Weaver, *J. Phys. A* **38**, 10433 (2005).
29. F. Sols, F. Guinea, A. H. Castro Neto, *Phys. Rev. Lett.* **99**, 166803 (2007).
30. The research was supported by Engineering and Physical Sciences Research Council (UK), the Royal Society, and Office of Naval Research. We are grateful to K. Ensslin, L. Eaves, M. Berry, L. Vandersypen, A. Morpurgo, A. Castro Neto, F. Guinea, and M. Fromhold for helpful discussions.
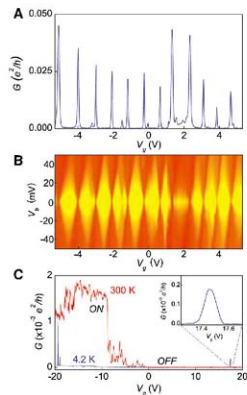
Fig. 4. Electron transport through nm-scale graphene devices. CB peaks (**A**) and diamonds (**B**) for a QD with an estimated size ~ 15 nm. (**C**) Electron transport through a conductance can be completely pinched-off even at room T. Fluctuations on the on state at room T are time dependent (noise excess). At low T, the on state exhibits much lower G, and the noise disappears. Occasional transmission resonances can also be seen as magnified in the inset.

Atomlike, Hollow-Core—Bound Molecular Orbitals of C60

Min Feng,[1] Jin Zhao,[1] Hrvoje Petek[1,2]*

The atomic electron orbitals that underlie molecular bonding originate from the central Coulomb potential of the atomic core. We used scanning tunneling microscopy and density functional theory to explore the relation between the nearly spherical shape and unoccupied electronic structure of buckminsterfullerene (C60) molecules adsorbed on copper surfaces. Besides the known π* antibonding molecular orbitals of the carbon-atom framework, above 3.5 electron volts we found atomlike orbitals bound to the core of the hollow C60 cage. These "superatom" states hybridize like the s and p orbitals of hydrogen and alkali atoms into diatomic molecule-like dimers and free-electron bands of one-dimensional wires and two-dimensional quantum wells in C60 aggregates. We attribute the superatom states to the central potential binding an electron to its screening charge, a property expected for hollow-shell molecules derived from layered materials.

C harge transport through organic materials occurs efficiently through covalently conjugated π molecular orbital networks, but not when molecules are bound by weak van der Waals forces (1, 2). Buckminsterfullerene (C60) molecular crystal is a typical strongly correlated Mott-Hubbard solid with a high effective electron mass. When intercalated by alkali atoms, C60 can be transformed into metallic, superconducting, insulating, and even magnetic phases (3). A low-temperature scanning tunneling microscopy (LT-STM) study of alkali atom–doped KxC60 monolayers (where x is the number of K atoms) on Au(111) surfaces found a doping-induced reentrant metal-insulator-metal phase transition involving previously unseen orientational ordering of C60 molecules. Wang et al. proposed that electron hopping between the π orbitals of the neighboring molecules creates anisotropic intermolecular interactions, which spontaneously drive the orientational ordering on the weakly interacting Au(111) surface (4, 5).

In an LT-STM study of C60 molecules on Cu surfaces, we found a completely different paradigm for the intermolecular electron delocalization involving isotropic interactions of atomlike electron orbitals centered on the nearly spherical C60 molecules. On Cu surfaces, the strong C60 molecule/substrate interactions hinder the spontaneous orientational ordering and concomitant electron delocalization through the intermolecular π bond hybridization (6). Partial occupation of the lowest unoccupied molecular orbital (LUMO) through chemisorption-induced charge transfer of up to two electrons per molecule facilitates charge transport through the overlayer, but poor intermolecular wave function overlap still hinders it within the overlayer (7, 8).

Thus, it is surprising that in LT-STM images (Fig. 1) above a bias voltage of 3.5 V, we found a transition from the π molecular orbital to a nearly free-electron (NFE)–like character, which is evident in the extensive, wave function delocalization within one-dimensional (1D) and 2D

assemblies of C60 molecules. We used LT-STM and density functional theory (DFT) to study the character of C60 molecular orbitals that give rise to these NFE properties. We found that the extensive hybridization involves atomlike orbitals bound by a central potential of the nearly spherical C60 shell. Because this potential is set up by the universal screening and surface dipole fields, we propose that these "superatom" electronic states are a common property of hollow molecules derived from wrapping or rolling of molecular sheets.

We investigated the electronic structure of C60 molecules on Cu(111) and partially oxidized Cu(110)-(2×1)-O surfaces (6, 9). By imaging chemisorbed C60 overlayers with constant-current scanning (Fig. 1, A to C) (10), we identified isolated-molecule, 1D wire, and partial 2D monolayer structures. We found the unoccupied states that mediate electron tunneling through C60 molecules by recording the distance-voltage (z-V) scans at different positions and numerically differentiating them to obtain dz/dV spectra such as that shown in Fig. 1K (11, 12). Finally, we explored the electronic nature

Fig. 1. Topographic (A to C) and d/dV (D to J) images of LDOS of a single C60 molecule, 1D quantum wire, and 2D quantum well. The quantum well is on a Cu(111) substrate, whereas all other images are for C60 molecules on 7.6 Å–wide bare Cu lines between oxide domains on an O/Cu(110) substrate (6). (K) dz/dV spectra identifying specific tunneling resonances in a single C60 molecule, a molecular dimer, and a C60 island. For potentials up to 3.2 V [(D) and (E)], the LDOS images show intramolecular structure characteristic of the π* orbitals localized on the C atom framework. The complete absence of molecular contrast in the LDOS images taken above 3.5 V for close-packed C60 aggregates is indicative of intermolecular wave function delocalization. The quantum-wire images show the transition from the localized molecular orbital LDOS of LUMO+2 at 2.8 to 3.2 V (E) to the delocalized one of s SAMO at 3.91 V (G). The complementary smooth contrast over both the C60 island (H) and the bare Cu(111) surface (I) regions reflects the NFE properties of SAMOs and

IP states. The arrows indicate the bright contrast at the edges and gaps of the quantum-wire antibonding band (G) and the dark contrast at the edge of the quantum-well bonding band (J).
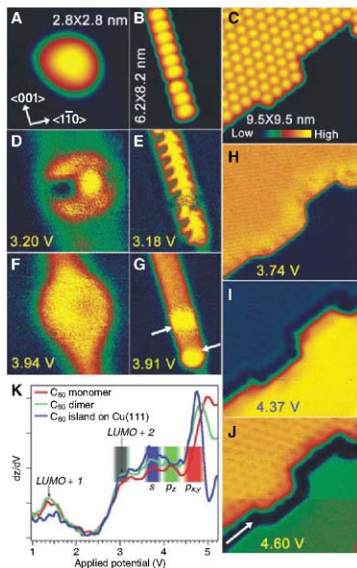
[1]Department of Physics and Astronomy and Petersen Institute of NanoScience and Engineering, University of Pittsburgh, Pittsburgh, PA 15260, USA. [2]Donostia International Physics Center, Paseo Manuel de Lardizabal 4, 20018 San Sebastian, Spain.

*To whom correspondence should be addressed. E-mail: petek@pitt.edu

of specific dz/dV resonances by imaging the local density of states (LDOS) with differential conductance (dI/dV) scanning (Fig. 1, D to J).

Our dz/dV spectroscopic survey from single C_{60} molecules to three molecular layers on O/Cu(110) and Cu(111) surfaces is consistent with previous assignments of the peaks at 1.5 and 2.8 to 3.2 V to the LUMO+1 and LUMO+2 states with π^* antibonding molecular orbital character [energies or bias potentials are referenced to the Fermi level (E_F)] (7, 8, 13, 14). The LUMO+2 images (Fig. 1, D and E) map through-molecule electron-tunneling pathways that depend on the surface-specific orientation of C_{60} molecules and the spatial distribution of their π^* molecular orbitals (7, 13, 15).

By contrast to these molecular orbital–resolved images, at potentials of 3.5 V and higher, the unoccupied states extend over the entire 1D and 2D assemblies (Fig. 1, G, H, and J). Such wave function delocalization gives similar contrast to the image of the complementary bare-metal region where, at 4.36 V, tunneling occurs through the intrinsic free-electron–like image potential (IP) state of the Cu(111) surface (Fig. 1I) (16).

Strong electron delocalization for molecular quantum wells on the C_{60}/Au(111) surface has also been reported by Zhu et al. (14). By angle- and time-resolved two-photon photoemission (2PP), Zhu et al. found a state at 3.7 eV with energy-momentum dispersion characterized by an effective mass only 1.4 times that of a free electron. On the basis of their DFT calculations performed with an atomic-orbital basis set, Zhu et al. attributed the delocalized state to the intermolecular hybridization of the LUMO+2 or LUMO+3 states.

To explain the marked change in the LT-STM images from the localized orbitals of LUMO+2 to the highly delocalized ones for the next higher-energy state, we performed plane-wave DFT calculations (17–20) for isolated fullerenes of different sizes, fullerenes with endohedral doping (21), and for C_{60} dimers. Here we have reproduced the well-known π^* antibonding LUMO orbitals of the sp² hybridized C atom network (13). In addition, we found diffuse orbitals, which, according to their angle-averaged wave functions in Fig. 2, are attached to the empty C_{60} core rather than the C atom shell and have nodes at the π orbital density maxima. On the basis of their radial nodes, these core-bound states have the principle quantum number $n = 3$, whereas the σ and π orbitals correspond to $n = 1$ and 2 (19). Because they evoke the spherical harmonic shape of s, p, and d symmetry atomic orbitals corresponding to the orbital angular momentum quantum numbers $l = 0, 1,$ and 2, we dub them the superatom molecular orbitals (SAMOs).

To confirm that the SAMOs calculated for free C_{60} molecules give rise to atomlike orbitals on Cu surfaces, we searched for their $l > 0$ counterparts and explored how SAMOs hybridize into delocalized bonds. Figure 3 shows dI/dV images taken at several energies above 3 V for a C_{60} molecule isolated in a Cu trough on the O/Cu(110) surface (6). The LDOS images corresponding to the three indicated positions of the dI/dz spectra of Fig. 1K (3.5 to 4.3, 4.6 to 4.9, and ≥5.0 V) exhibit distinct monopolar (quantum number $m = 0$), dipolar ($m = 1$), and quadrupolar ($m = 2$) symmetry, corresponding to their dominant s, p, and d orbital character (m is the surface projection of l). On the basis of the $m = 0$ symmetry and energy, we attribute the 3.70 and 4.15 V images to the s and p_z SAMOs. The s SAMO has a characteristic dark spot (diminished LDOS) in registry with the topmost pentagon of the C_{60} molecule, whereas the p_z SAMO is more compact. The broken symmetry at the surface hybridizes the s and p_z SAMOs into a bonding (3.70 V) and an antibonding (4.15 V) pair. The anisotropy of the O/Cu(110) surface lifts the degeneracy of the $m = 1$ SAMOs; the repulsion by the oxide domains along the ⟨1̄10⟩ azimuth destabilizes the p_y SAMO with respect to the p_x SAMO. Finally, based on its appearance, we assign the 5.0 V image to hybridization of the p_y and d_{yz} SAMOs. Because several p-d hybrid orbitals are likely to fall into a narrow energy range, the assignment of d SAMOs is uncertain.

Having confirmed the existence of hydrogen or alkali atomlike SAMOs, we then explored their hybridization into delocalized states of super-atomic matter. Figure 4 presents the calculated wave functions and observed images of s and p SAMOs hybridizing into σ and π symmetry bond-
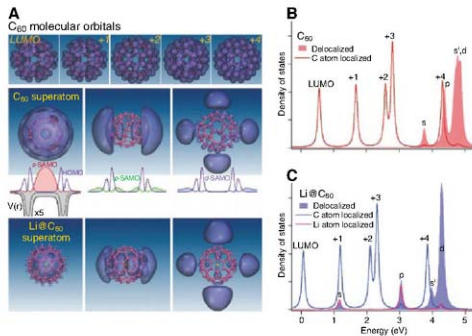
Fig. 2. (**A**) Calculated LUMO to LUMO+4 and s, p, and d SAMO wave functions for isolated C_{60} and Li@C_{60} molecules. The angle-averaged occupied-state wave functions find the σ and π electron probability at, as well as above and below, the C atom shell plane, whereas the SAMO probability density is centered on the hollow molecular core. The SAMO radial nodes appear at the π electron density maxima, as required by the $n = 3$ principle quantum number and the Pauli exclusion principle. The black line gives the angularly averaged DFT density, $V(r)$, showing the deep well of the C atom shell and the shallow screening potential of the hollow C_{60} core (expanded five times). (**B** and **C**) C and Li atom localized and delocalized DOS for C_{60} and Li@C_{60}. Charge transfer of the 2s electron from Li [yellow in center of C_{60} for Li@C_{60} in (A)] to the LUMO of C_{60} causes the reference energy to shift between (B) and (C). Hybridization with the Li atomic orbitals lowers the energy of SAMOs with respect to the undoped case. s′ refers to an $n = 4$ state.

Fig. 3. dI/dV mapping of an individual C_{60} molecule's SAMOs. As the measurement voltage is increased, the dI/dV images of a single C_{60} molecule evolve from the π^* molecular orbital character of LUMO+2 to the core-centered s, p_z, p_x, p_y, and d_{yz} SAMO character. The characteristic dark spot on the s SAMO occurs above the top C_{60} pentagon. Tunneling through the image potential states of the substrate is responsible for the bright contrast above the metal and the oxide domains.
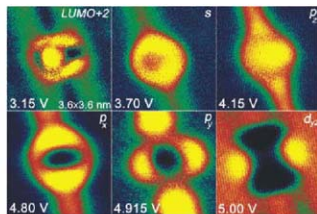
ing and antibonding orbitals of superatom dimers. The molecular orbitals derived from SAMOs, like those of an H_2 molecule, can be classified by their nodal pattern and is orthogonal to the molecular plane. The 3.5 to 4.3 V region encompasses the bonding and antibonding hybridization of the s and p_z SAMOs. The 4.60 to 4.95 V region shows a progression from the σ bonding of the p_x SAMOs to the π^* antibonding of the p_x and p_y SAMOs. Above 4.90 V, the dimer orbitals are more diffuse and may have additional d SAMO character. Though obscured by congestion and broadening, the comparison between the dz/dV spectra of the monomer and the dimer (Fig. 1K) suggests that the hybridization-induced shifts are small (<0.2 V). By contrast to SAMOs, the LUMO+2 dimer images in Fig. 4 show little evidence of the intermolecular hybridization that could give rise to NFE character.

In larger aggregates, SAMOs hybridize into 1D quantum wires (Fig. 1, H and J). Electron delocalization within these extended structures is responsible for the vanishing molecular contrast and other telling features. Whereas a SAMO–derived images show uniform contrast, the $m = 1$ quantum well on Cu(111) (4.60 V) (Fig. 1J), reflecting its origin in the hybridization of the degenerate p_x and p_y SAMOs, has nodes centered on C_{60} molecules. Furthermore, near the edges of 1D chains and 2D islands (Fig. 1, G and J), the energy-dependent bonding and antibonding hybridization gives rise to distinct dark or bright boundary states (22).

Having established that above 3.5 V SAMOs dominate the STM dI/dV images of single and aggregated C_{60} molecules, we next considered their physical origin. Emergence of the atomic-orbital character in a molecular system requires a central potential. For example, atomlike molecular Rydberg states exist because weakly bound electrons interacting with a molecular ion through the long-range Coulomb potential essentially do not penetrate into the molecular core. The SAMOs, however, are distinct from Rydberg states because their probability density resides substantially—and, for the s SAMO, dominantly—within the hollow C_{60} core. Considering the C_{60} molecule as a hollow dielectric shell whose screening properties are described by the angle-averaged DFT potential in Fig. 2, we can reproduce the radial shapes but not the exact energies of the LUMO and SAMO wave functions.

To explore the origin of SAMOs, we solved the Schrödinger equation using the form of the angle-averaged DFT potential but varying the relative core and shell contributions. These calculations show that the shell potential dominantly gives rise to the $n = 2$ HOMO (highest occupied molecular orbital) and LUMO orbitals, whereas the nearly constant attractive screening potential extending 0.25 nm from the molecular center is essential for the existence of the $n = 3$ SAMOs (19, 23). The energy and shape of SAMOs also depend on the C atom shell potential, which locates the radial nodes at the π orbital density maxima.

The origin of the central potential within the C_{60} core can be traced to the screening of an electron charge through the short-range exchange and correlation and the long-range Coulomb interactions. At a solid/vacuum interface, these many-body interactions give rise to a series of IP states that converge to the vacuum level and have NFE properties parallel to the surface (24). For 2D molecular sheets such as graphene and hexagonal BN, IP states should exist on both sides and hybridize into symmetric and antisymmetric linear combinations (25, 26). Stacking molecular sheets to form quasi-2D solids causes IP states to hybridize into interlayer bands characterized by a probability-density peak between and NFE properties parallel to the sheets (25, 26).

Topological distortion of molecular sheets (that is, wrapping or rolling them into 1D or 0D molecules) also affects the bilateral IP states; the favorable screening of the concave surface stabilizes the interior IP state with respect to the exterior one. Also, a dipole directed along the surface normal that is created by the curvature-induced hybridization of the σ and π molecular states contributes an additional attractive potential (26–28). The resulting internal potential gives rise to SAMOs, which can be thought of as 0D interlayer states, whereas the external IP potential is responsible for the more weakly bound Rydberg states (29). Viewed as a consequence of the universal surface screening and dipole potentials, SAMOs emerge as a general property acquired by rolling and wrapping 2D layered materials into 1D nanotubes and 0D hollow molecules.

The 2PP band dispersions (14) and LT-STM images of delocalized wave functions together demonstrate that SAMOs impart metal-like conductivity to the excited anion states of self-assembled C_{60} structures. Because these properties derive from the topological modification and assembly, we postulate that SAMOs can impart previously unseen electronic and optical properties to other layered materials. For example, the conduction bands of the hexagonal BN and BN nanotubes have NFE interlayer-state character (26). The SAMO properties also can be tuned through the cage size or intercalation. In Fig. 2, we present the calculated SAMO orbitals and the C and Li atom localized and delocalized density of states (DOS) for Li@C_{60}, which illustrate that an endohedral Li atom in the center of a C_{60} molecule stabilizes the s SAMO to 1.2 eV above E_F. In other calculations on single-wall carbon nanotubes, the internal doping with a positive charge was shown to stabilize an interlayer-derived NFE state by more than 3 eV to below E_F (27). Because the image and dipole potentials are universal properties of the curved molecular sheets, the NFE states analogous to SAMOs should be a general property of hollow molecules such as fullerenes.

References and Notes
1. V. Coropceanu et al., Chem. Rev. 107, 926 (2007).
2. F. S. Tautz, Prog. Surf. Sci. 82, 479 (2007).
3. O. Gunnarsson, Rev. Mod. Phys. 69, 575 (1997).
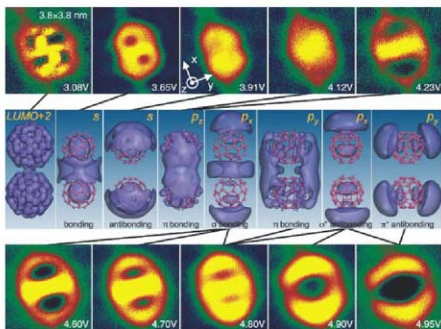4. A. Wachowiak et al., Science 310, 468 (2005).

Fig. 4. Hybridization of the superatom orbitals of individual C_{60} molecules into the corresponding molecular orbitals of C_{60} dimers was investigated by DFT calculations and dI/dV imaging. The calculated orbitals are for isolated dimers, whereas the experimental LDOS images are for dimers on Cu troughs. The H atom–like s and p SAMOs combine to form σ and π symmetry bonding and antibonding molecular orbitals resembling those of an H_2 molecule. The correspondence between the calculated probability densities and measured LDOS is proposed on the basis of the observed nodal patterns. Overlap of several bands complicates assignments, particularly at higher measurement potentials. The dimer images show how the SAMO character evolves into the NFE properties of superatom matter. By contrast to SAMO images, LUMO+2 does not show substantial hybridization-induced changes in the LDOS.

5. Y. Wang et al., Phys. Rev. Lett. **99**, 086402 (2007).
6. M. Feng, J. Lee, J. Zhao, J. T. Yates, H. Petek, J. Am. Chem. Soc. **129**, 12394 (2007).
7. T. Hashizume et al., Phys. Rev. Lett. **71**, 2959 (1993).
8. K. D. Tsuei, P. D. Johnson, Solid State Commun. **101**, 337 (1997).
9. Experiments were carried out in an ultrahigh vacuum (base pressure <10⁻¹⁰ mbar) LT-STM. Samples were prepared as described in (6). The templates for the C_{60} monomer, dimer, and wire were occasional 7.6 Å–wide unoxidized rows of single Cu atoms cutting nearly defect-free Cu(110)-(2×1)-O domains. The intermolecular distance within a wire of 10.7 ± 0.2 Å is ~7% larger than the distance in a C_{60} solid (10.05 Å). All measurements were performed at 77.3 K.
10. dI/dV images were acquired by positioning the scanning tunneling microscope tip at each point under constant current (engaged-loop conditions) and then measuring the dI/dV signal with lock-in detection (tunneling current driven at 267 Hz with 18-mV modulation of the junction bias under open-loop conditions).
11. D. B. Dougherty, P. Maksymovych, J. Lee, J. T. Yates Jr., Phys. Rev. Lett. **97**, 236806 (2006).
12. STM topography images were acquired in constant-current mode (0.1-nA current and 0.45-V sample potential referenced to the tip). Tunneling spectroscopy of the unoccupied states was performed by measuring the constant-current z-V characteristics as described in (11). Tunneling resonances are identified by the numerical differentiation of the z-V data. The similarity between dI/dV spectra of C_{60} on the Cu/Cu(110) and

C_{60}/Cu(110)-(2×1)-O surfaces below 5 V indicates weak perturbation of the electronic structure by the oxide.
13. X. Lu, M. Grobis, K. H. Khoo, S. G. Louie, M. F. Crommie, Phys. Rev. B **70**, 115418 (2004).
14. X. Y. Zhu et al., Phys. Rev. B **74**, 241401 (2006).
15. DFT calculations predict LUMO+2 and LUMO+3 to be separated by 0.17 eV and to have similar orbital distributions. Experimentally, in the 2.8 to 3.2 V range, dI/dV spectra and dI/dV imaging do not show evidence for two distinct states. Whereas we attribute our experimental observations to LUMO+2, the observed DOS may also have contributions from LUMO+3.
16. D. B. Dougherty et al., Phys. Rev. B **76**, 125428 (2007).
17. J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
18. Using the generalized gradient approximation to DFT with the Perdew-Becke-Ernzerhof functional, we calculated the electronic structure of isolated C_{60} molecules (18). The 30-by-30-by-30 Å unit cell was used to avoid the interaction between C_{60} molecules. The energy cutoff for the plane waves was 400 eV. The results were checked using localized density approximation with the Ceperley and Alder functional, and the difference was negligible (19). SAMO DOS represents the difference between the total DOS and that projected onto the atomic orbitals on C and Li atoms. Calculating the localized and delocalized states on an equal footing is best accomplished with a plane-wave basis set, as done in (20) for a C_{60}

solid. To characterize the diffuse states with an atomic basis requires an appropriately large basis set.
21. H. Shinohara, Rep. Prog. Phys. **63**, 843 (2000).
22. J. N. Crain, D. T. Pierce, Science **307**, 703 (2005).
23. P. Delaney, J. C. Greer, Appl. Phys. Lett. **84**, 431 (2004).
24. P. M. Echenique, J. B. Pendry, J. Phys. C Solid State Phys. **11**, 2065 (1978).
25. M. Posternak, A. Baldereschi, A. J. Freeman, E. Wimmer, M. Weinert, Phys. Rev. Lett. **50**, 761 (1983).
26. X. Blase, A. Rubio, S. G. Louie, M. L. Cohen, Phys. Rev. B **51**, 6868 (1995).
27. E. R. Margine, V. H. Crespi, Phys. Rev. Lett. **96**, 196803 (2006).
28. Y. Miyamoto, S. Saito, D. Tománek, Phys. Rev. B **65**, 041402 (2001).
29. M. Boyle et al., Phys. Rev. Lett. **87**, 273401 (2001).
30. We thank J. Lee and J. T. Yates for their contributions to templated C_{60} growth, sharing their expertise, and provisioning the LT-STM. We also thank R. W. Field, J. Yang, and D. Pratt for discussions. Grants from the U.S. Department of Energy (DOE DE-FG02-03ER15434), the Petroleum Research Fund (44158-ACS), and Ikerbasque supported this research. Calculations were performed in the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory, a user facility sponsored by the DOE Office of Biological and Environmental Research.

30 January 2008; accepted 14 March 2008
10.1126/science.1155866

The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes

Maureen E. Hillenmeyer,[1,2] Eula Fung,[1] Jan Wildenhain,[3]* Sarah E. Pierce,[1,4] Shawn Hoon,[1,4] William Lee,[1,4] Michael Proctor,[1] Robert P. St.Onge,[1] Mike Tyers,[3,5,]* Daphne Koller,[6] Russ B. Altman,[2,4] Ronald W. Davis,[2,4] Corey Nislow,[5,7,8] Guri Giaever[5,8,9]†

Genetics aims to understand the relation between genotype and phenotype. However, because complete deletion of most genes (~80%) has no obvious phenotypic consequence in rich medium, it is difficult to study their functions. To uncover phenotypes for this nonessential fraction of the genome, we performed 1144 chemical genomic assays on the yeast whole-genome heterozygous and homozygous deletion collections and quantified the growth fitness of each deletion strain in the presence of chemical and environmental stress conditions. We found that 97% of gene deletions exhibited a measurable growth phenotype, suggesting that nearly all genes are essential for optimal growth in at least one condition.

Small molecules are potent probes to help understand cellular physiology [for review, see (1)]. The emergent field of chemical genomics promises that, by understanding the relations between small molecules and genes on a systems level, we might understand genomic responses to small molecule perturbants. We show that the global response of all protein-coding gene deletions tested with a panel of several hundred perturbations yields insight into gene dispensability, multidrug resistance, and gene functions within the *Saccharomyces cerevisiae* cell.

The diploid yeast deletion collections comprise ~6000 heterozygous gene deletion strains and ~5000 homozygous gene deletions (~1000 genes are essential) (2, 3). We tested the growth responses of these cells to over 400 small molecules and diverse environmental stresses.

Surveying a large swath of ecological space allowed us to identify genes required for growth in each tested condition. Essential genes are a potential source of new drug targets (4), whereas nonessential genes have been proposed to contribute to genetic robustness (via compensation by redundant pathways) (5, 6) or to be required for growth in particular conditions (7). Our results provide an experimental framework to test these hypotheses. We also identified previously unknown genes that function in multidrug resistance (MDR), that is, those genes required for growth in the presence of multiple drugs.

We screened small molecules from diverse sources and libraries, including drugs approved by the World Health Organization and the U.S. Food and Drug Administration, well-characterized chemical probes, and compounds with uncertain

biological activity (tables S1 and S2). The structural diversity of these compounds is comparable to that of approved drugs (fig. S1). We also assayed the effects of various environmental treatments and stresses (for example, depletion of amino acids or vitamins). We performed 726 treatment experiments in each of the heterozygous deletion strains and 418 separate experiments in each of the homozygous strains, for a total of more than 6 million single-gene measurements. These sets include some repeated experiments in which drug dose or exposure time was varied. Collapsing such repeats yielded a total of 354 unique conditions for the heterozygous collection and 178 for the homozygous collection (124 of which were tested against both collections). A gene deletion strain was defined as sensitive to a treatment if it showed a growth defect in the treatment relative to its growth in control (no drug) conditions. We defined significant sensi-

[1]Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304, USA. [2]Program in Biomedical Informatics, Stanford University, Palo Alto, CA 94305, USA. [3]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario M5G1X5, Canada. [4]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S1A8, Canada. [6]Department of Computer Science, Stanford University, Palo Alto, CA 94305, USA. [7]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5S3E1, Canada. [8]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S3E1, Canada. [9]Department of Pharmaceutical Sciences, University of Toronto, Toronto, Ontario M5S3M2, Canada.

*Present address: Wellcome Trust Centre for Cell Biology, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK.
†To whom correspondence should be addressed. E-mail: guri.giaever@utoronto.ca

tivity and corrected for multiple comparisons by controlling the false discovery rate (FDR) to ~0.1 for genes exhibiting any phenotype (8) (figs. S2 and S3).

Previous studies revealed that 34% of homozygous deletion strains display a distinct phenotype (19% lethality and 15% fitness defect) when grown in rich medium (2, 3, 9). Three percent of heterozygous strains display a fitness defect (9). One interpretation of these observations is that the majority of the yeast genome is dispensable for growth. However, it is unlikely that yeast encounters such ideal conditions outside of the

laboratory. In our experiments, nearly all of the deletion strains manifested a phenotype in one or more conditions (Fig. 1).

In the experiments using homozygous deletion strains, only 416 strains (containing deletions of 7% of the genome) did not manifest a phenotype that was different from the no-drug control phenotype (hereafter "exhibited no phenotype"). Because more than half of these 7% of the genome are either of unknown function or designated as dubious (10), it is likely the majority do not encode proteins. Of the heterozygous deletion strain experiments, 2049 (34%

of the genome) exhibited no phenotype. Nearly a third of these are either of unknown function or designated as dubious (10) and are also unlikely to be protein-coding. Only 205 strains (3%) failed to exhibit a phenotype under either homozygous or heterozygous conditions. Thus, considering all ~6000 yeast genes, 97% manifested a significant fitness defect when deleted. In a control analysis of rich-medium experiments, only about 10% of strains manifested a false positive fitness defect [FDR of ~0.1 (Fig. 1A and figs. S2 and S3)]. We further tested the small set of 205 deletion strains that showed no
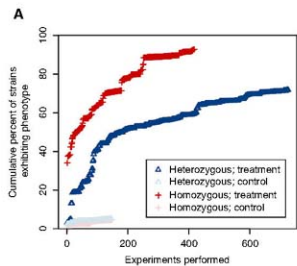
Fig. 1. Fraction of genome required for optimal growth under experimental conditions. (**A**) Percent of gene deletion strains that exhibit significant sensitivity in at least one treatment as a function of number of experiments performed, ordered by date. We used a significance threshold (z score $P < 1 \times 10^{-5}$ and $P < 1 \times 10^{-6}$ for homozygous and heterozygous experiments, respectively) that limited the FDR of genes exhibiting any phenotype to ~0.1 (figs. S2 and S3). Treatment experiments measure the growth of the deletion strains in a drug or altered environmental conditions; control experiments measure growth of the same deletion strains in no-drug medium (8). The percent of strains exhibiting a phenotype begins at the percentage previously observed in rich medium (3% for heterozygotes and 34% for homozygotes). (**B**) Percentage of yeast genes with a phenotype under particular conditions: 19% are essential genes, 15% exhibited a growth defect as homozygous deletions in rich medium, and 63% exhibited a phenotype as either homozygous or heterozygous deletions under particular conditions in this study.
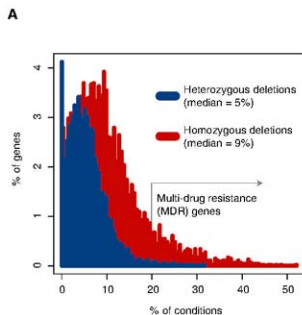
Fig. 2. Genes required for optimal growth in multiple conditions. (**A**) The percent of deletion strains inhibited by the given percent of unique conditions at $P < 0.01$ (z score). Most strains are perturbed by multiple distinct conditions. Genes whose deletion strains showed sensitivity to at least 20% of the unique small molecules are defined as MDR genes. (**B**) Enriched functions of the homozygous MDR genes. Shown are the important GO biological processes, molecular functions, and cellular components, the hypergeometric P value, and the locations of the processes in the cell. AP, adaptor protein; ESCRT, endosomal sorting complex required for transport.
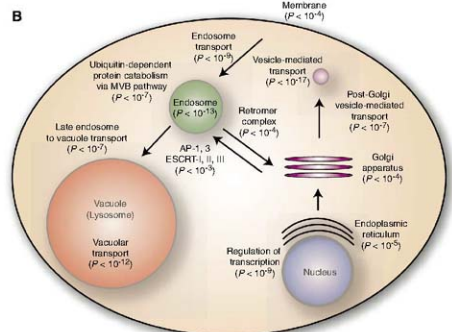
growth defect: 62 deletions do not appear to encode proteins (10), and 155 are expressed at low levels or not at all (11). This gene set was not enriched for duplicated genes (12), as might have been predicted by the redundancy hypothesis (5–7), nor were duplicated gene deletion strains sensitive in a greater or fewer number of conditions on average (fig. S5).

Certain deletion strains were sensitive to multiple drug treatments (Fig. 2A) (13–15), and therefore the corresponding deleted genes appear to be required for resistance to diverse perturbations. We therefore refer to them as multidrug resistance (MDR) genes, with "multiple" defined as greater than 20% of all unique treatments. We predicted that MDR genes in our data would include well-characterized pleiotropic efflux pumps such as PDR5 (human homolog is MDR1) and their regulators such as the transcription factor PDR1 (16). These transporters do score as MDR in our assay, but several hundred additional genes behaved similarly: In total, 51 genes were defined as MDR as heterozygous deletion strains, and 510 were defined as MDR as homozygous deletion strains (Fig. 2A). These genes are highly enriched for certain Gene Ontology functions (17), particularly endosome transport, vacuolar degradation, and transcription; together these cellular processes compose a coherent system (Fig. 2B). This coordinated system of endocytosis and vacuolar or lysosomal degradation, conserved from yeast to humans, is a mechanism whereby the cell regulates transmembrane proteins, including the transporters PDR5, FUR4, and TAT1 and signaling receptors (18). Subsets of these highly conserved pathways have been previously associated with MDR in yeast (13–15, 19). These results are consistent with recent findings in mammalian cells that MDR is correlated with changes in intracellular trafficking (20, 21), although the exact contribution of these pathways to drug resistance is not known (22–25). The most frequently observed MDR gene was IRS4, and although it has not previously been associated with any drug response, we found that this gene conferred resistance to 55% of the unique compounds. IRS4 regulates phosphoinositides (PIs), well-conserved second messengers that regulate vesicular trafficking, membrane transporters, and membrane lipid composition (18). These functions largely encompassed the remaining MDR genes' functions (Fig. 2B), suggesting that IRS4 signaling may coordinate MDR-specific pathways. To address this, we examined the MDR genes that conferred resistance to the same compounds as IRS4; these genes were enriched for vesicle transport function (hypergeometric $P < 1 \times 10^{-20}$), whereas those genes not related to IRS4 were not enriched. Additional interesting MDR genes included nine genes involved in aromatic amino acid biosynthesis (TRP1, TRP2, TRP3, TRP4, TRP5, ARO1, ARO2, ARO7, and TKL1) and GCN4, a master transcriptional regulator of amino acid biosynthesis (15). The sensi-

tivity profiles of these strains were nearly identical to one another, and the strains were sensitive to nearly 30% of all tested compounds. Lastly, although transcription factors (TFs) are underrepresented in the set of essential genes in rich media (~3% of

TFs were essential) (26), we found that 16 TFs (~10%) conferred MDR. Our compendium of conditions under which TFs are required for growth suggests experimental conditions to better characterize these proteins.
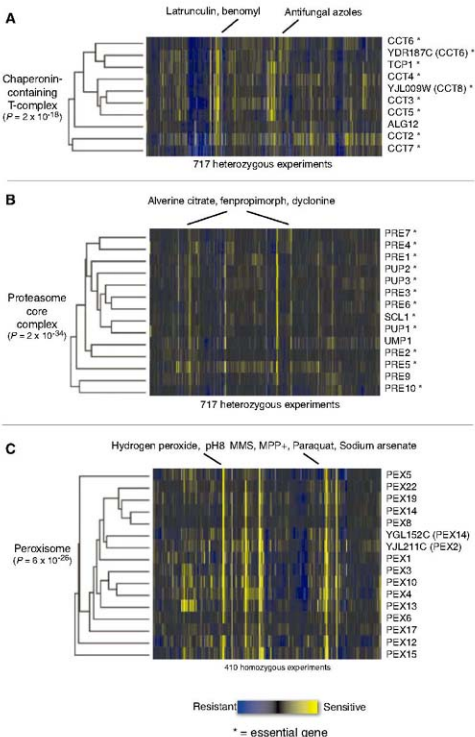
Fig. 3. Gene clusters with similar co-fitness profiles and similar biological functions, extracted from two-way hierarchical clustering on the complete data set (using all genes and all experiments). Each cell in the matrix is a sensitivity score of the deletion strain in the experiment: yellow indicates that the strain was sensitive; blue indicates resistance. Essential genes are marked with asterisks, and open reading frames (ORFs) that overlap a dubious ORF are in parentheses. (A) CCT complex genes, mediators of cytoskeletal assembly, cluster as heterozygous deletion strains primarily because of sensitivity to the cytoskeletal disrupting agents latrunculin and benomyl. (B) Genes of the proteasome core complex cluster in the heterozygous deletion assays because of sensitivity to three structurally similar compounds that target the lipid synthesis pathway. (C) Peroxisomal genes clustered as homozygous deletions because of sensitivity to hydrogen peroxide, high pH, and oxidative stress-inducing conditions.

We defined the co-fitness between any two genes as the correlation of their fitness profiles across all experiments (19). We clustered genes by this metric and found that those that are co-fit often clustered by known function, having shared Gene Ontology biological processes and molecular functions (17). Three examples of function-enriched clusters were extracted from genomewide clustering, and they serve to validate co-fitness as an informative metric that provides biological insight into the function of these genes. One heterozygous cluster (Fig. 3A) comprised all eight chaperonin-containing T (CCT) complex genes: *TCP1*, *CCT2*, *CCT3*, *CCT4*, *CCT5*, *CCT6* (and its overlapping gene *YDR187C*), *CCT7*, and *YJL009W*, which overlaps *CCT8*. The CCT cytoskeletal-folding complex is essential and conserved between yeast and human. These strains are especially sensitive to the cytoskeletal poisons latrunculin and benomyl, illustrating cases where drugs and gene deletions act synergistically in an essential pathway to produce a phenotype. Another heterozygous deletion strain cluster included 13 of the 14 proteasome core complex genes and *UMP1*, a chaperone required for maturation of that complex (Fig. 3B). These strains were sensitive to alverine citrate, fenpropimorph, dyclonine, and dihydromotuporamine C, all of which share a nitrogen-containing structural motif and may target the lipid synthesis pathway (4). This observation suggests a possible relation between lipid synthesis and proteasome-mediated degradation. The cluster of homozygous strains shown in Fig. 3C contains 15 genes encoding components of the peroxisome, an organelle conserved from yeast to human that is responsible for breakdown of peroxides and other metabolites (27). Not surprisingly, the deletion strains were sensitive to hydrogen peroxide, presumably because the defective peroxisome could not properly metabolize it. The peroxisome requires a low pH, and the deletion strains were also sensitive to high pH. There is also growing evidence for peroxisomal involvement in managing oxidative stress (27); interestingly, the deletion strains were sensitive in the presence of chemicals that induce oxidative stress, such as MPP+ (1-methyl-4-phenylpyridinium), paraquat, and sodium arsenate.

Our finding that nearly all genes in yeast are required for wild-type growth in at least one experimental condition addresses the debate concerning the purpose of nonessential genes. The definition of essentiality is controversial, but under laboratory conditions over 80% of the genome is dispensable for life. We have found conditions that render the remaining genes essential for optimal growth in some condition and show that genetic redundancy is therefore limited in providing tolerance of the tested conditions (5–7). The chemical-genetic interactions defined here are complementary to genetic interactions (28) and should improve the resolution of genetic and chemical-genetic interaction maps, with applications such as predicting drug activities and synergies (29, 30).

References and Notes
1. S. L. Schreiber, *Nat. Chem. Biol.* **1**, 64 (2005).
2. E. A. Winzeler *et al.*, *Science* **285**, 901 (1999).
3. G. Giaever *et al.*, *Nature* **418**, 387 (2002).
4. G. Giaever *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 793 (2004).
5. G. Gu *et al.*, *Nature* **421**, 63 (2003).
6. D. Deutscher, I. Meilijson, M. Kupiec, E. Ruppin, *Nat. Genet.* **38**, 993 (2006).
7. B. Papp, C. Pal, L. D. Hurst, *Nature* **429**, 661 (2004).
8. Materials and methods are available as supporting material on Science Online and at http://chemogenomics. stanford.edu/supplement/global.
9. A. M. Deutschbauer *et al.*, *Genetics* **169**, 1915 (2005).
10. www.yeastgenome.org/
11. L. David *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5320 (2006).
12. M. Kellis, B. W. Birren, E. S. Lander, *Nature* **428**, 617 (2004).
13. A. B. Parsons *et al.*, *Nat. Biotechnol.* **22**, 62 (2004).
14. A. B. Parsons *et al.*, *Cell* **126**, 611 (2006).
15. R. C. Fry, T. J. Begley, L. D. Samson, *Annu. Rev. Microbiol.* **59**, 357 (2005).
16. H. Jungwirth, K. Kuchler, *FEBS Lett.* **580**, 1131 (2006).
17. E. I. Boyle *et al.*, *Bioinformatics* **20**, 3710 (2004).
18. D. J. Katzmann, G. Odorizzi, S. D. Emr, *Nat. Rev. Mol. Cell Biol.* **3**, 893 (2002).
19. A. M. Dudley, M. A. Janse, A. Tanay, R. Shamir, G. McDonald Church, *Mol. Systems Biol.* **1** (2005); www.nature.com/msb/journal/v1/n1/full/msb4100004.html.
20. X. J. Liang, S. Mukherjee, D. W. Shen, F. R. Maxfield, M. M. Gottesman, *Cancer Res.* **66**, 2346 (2006).
21. A. Rajagopal, S. M. Simon, *Mol. Biol. Cell* **14**, 3389 (2003).
22. R. D. Bagshaw, D. J. Mahuran, J. W. Callahan, *Mol. Cell. Proteomics* **4**, 133 (2005).
23. Y. Gong, M. Duwvari, J. P. Krise, *J. Biol. Chem.* **278**, 50234 (2003).
24. R. Egner, K. Kuchler, *FEBS Lett.* **378**, 177 (1996).
25. J. Ferte, *Eur. J. Biochem.* **267**, 277 (2000).
26. G. Chua, M. D. Robinson, Q. Morris, T. R. Hughes, *Curr. Opin. Microbiol.* **7**, 638 (2004).
27. M. Schrader, H. D. Fahimi, *Histochem. Cell Biol.* **122**, 383 (2004).
28. A. H. Y. Tong *et al.*, *Science* **303**, 808 (2004).
29. J. R. Sharom, D. S. Bellows, M. Tyers, *Curr. Opin. Chem. Biol.* **8**, 81 (2004).
30. P. E. Blower *et al.*, *Pharmacogenomics J.* **2**, 259 (2002).
31. We thank J. Rine, C. Boone, K. Kuchler, E. Ericson, and members of the Giaever and Nislow labs for comments. Supported by grants from the National Human Genome Research Institute (to R.W.D, C.N., and G.G.), the NSF (to M.E.H. and D.K.), the Canadian Institute for Health Research (no. 81340 to G.G. and C.N.), and the Canadian Research Chair in Genomics (to G.G.).

Wnt5a Control of Cell Polarity and Directional Movement by Polarized Redistribution of Adhesion Receptors

Eric S. Witze,[1] Elizabeth S. Litman,[1,2] Gretchen M. Argast,[1] Randall T. Moon,[3,2] Natalie G. Ahn[1,2]*

Mechanisms by which Wnt pathways integrate the organization of receptors, organelles, and cytoskeletal proteins to confer cell polarity and directional cell movement are incompletely understood. We show that acute responses to Wnt5a involve recruitment of actin, myosin IIB, Frizzled 3, and melanoma cell adhesion molecule into an intracellular structure in a melanoma cell line. In the presence of a chemokine gradient, this Wnt-mediated receptor–actin–myosin polarity (W-RAMP) structure accumulates asymmetrically at the cell periphery, where it triggers membrane contractility and nuclear movement in the direction of membrane retraction. The process requires endosome trafficking, is associated with multivesicular bodies, and is regulated by Wnt5a through the small guanosine triphosphatases Rab4 and RhoB. Thus, cell-autonomous mechanisms allow Wnt5a to control cell orientation, polarity, and directional movement in response to positional cues from chemokine gradients.

Wnt signaling controls cell polarity and directional cell movement in developmental systems, as well as cell invasion in certain cancers. Features shared between noncanonical Wnt pathways include recruitment of Frizzled (Fz) receptors to the posterior end of cells, and asymmetric distribution of atypical cell adhesion molecules, often associated with Fz

(1, 2). Thus, receptors redistribute, in response to Wnt, to define an axis of asymmetry. In developmental systems, these processes can be regulated by interactions with adjacent cells, which confer orientation with respect to surrounding tissues (1, 3). For example, during endoderm specification in *Caenorhabditis elegans*, the division plane in the four-cell blastomere is determined by a positional Wnt signal from a nearby P₂ cell (4). In contrast, Wnt pathway mutations in *C. elegans* that disrupt neuronal cell migration and polarity can be rescued by Wnt overexpression without requiring a localized source of ligand (5, 6). This suggests that directional presentation of Wnt to cells is not always needed for

[1]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA. [2]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA. [3]Department of Pharmacology and Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: natalie.ahn@colorado.edu

cell polarization, rather that Wnt ligands act permissively to allow cells to respond to environmental cues for longitudinal guidance and directional movement. In such cases, intracellular events responding directly to Wnt may be difficult to identify and distinguish from responses to cell-cell contact or paracrine signaling.

To address this, we examined acute responses to Wnt5a in dispersed cells. Wnt5a expression correlates with high-grade, invasive human melanomas and promotes invasiveness in melanoma cell lines [7]. Treatment of WM239A melanoma cells with purified Wnt ligands showed expected signaling selectivity, where Wnt3a enhanced β-catenin–dependent transcription and nuclear translocation, while Wnt5a enhanced protein kinase C (PKC) autophosphorylation with little effect on β-catenin–dependent signaling (fig. S1) [8].

WM239A cells treated with Wnt5a were immunostained for melanoma cell adhesion molecule (MCAM, also known as MUC18 and CD146), an immunoglobulin G–family cell adhesion receptor implicated in melanoma tumorigenesis and metastasis [9]. In untreated cells, MCAM was distributed uniformly. However, treatment of cells with Wnt5a for 30 min led to redistribution of MCAM into a polarized structure, which we call the Wnt5a-mediated receptor–actin–myosin polarity (W-RAMP) structure, for reasons described below (Fig. 1, A and B). The W-RAMP structure was also observed in some untreated cells, but this was suppressed by depleting Wnt5a with RNA interference (RNAi) (Fig. 1B), which indicated that the basal signal results from autocrine responses to endogenous ligand. Similar W-RAMP structures were observed in WM1789 melanoma cells, as well as in xenograft tumors derived from WM239A cells (Fig. 1A and fig. S2A). In contrast, Wnt3a did not induce the W-RAMP structure over that of the controls did (Fig. 1B).

In order to address its relation to cell polarization and orientation, the position of the W-RAMP structure was examined relative to Golgi, which in polarized motile cells typically localizes to the anterior end. To provide a localized cue for positional orientation, melanoma cells were exposed to a gradient of CXCL12, a chemokine that signals through the CXCR4 receptor to promote cell invasion [10, 11]. WM239A cells were placed in a chemotaxis chamber in the presence or absence of Wnt5a and with CXCL12 in the chemoattractant source chamber. After 30 min, cells were immunostained for MCAM and a Golgi marker.

Three patterns were observed: (i) W-RAMP structures were located distal to Golgi with respect to the nucleus, consistent with a posterior location; (ii) W-RAMP structures were proximal to Golgi; or (iii) Golgi were not clearly polarized. Wnt5a treatment increased the percentage of cells with Golgi positioned distal to the W-RAMP structures and decreased the percentages of other patterns (Fig. 1C). This indicates that Wnt5a enhances cell polarization, directing MCAM to the distal end. It is noteworthy that the response

completely depended on the presence of chemokine, because without CXCL12, distal juxtapositioning between MCAM and Golgi did not increase and, in fact, slightly decreased following Wnt5a treatment (Fig. 1C). Thus, Wnt5a promotes cell polarity in a manner that requires the CXCL12 gradient.

We examined the orientation of Golgi with respect to the CXCL12 gradient source (Fig. 1D). Wnt5a increased the number of cells in which Golgi were pointed either toward or away from the CXCL12 gradient and decreased the number of cells in which Golgi were not clearly polarized (Fig. 1D). However, Wnt5a did not substantially affect proximity of the W-RAMP structure relative to the gradient source, and cells became polarized with Golgi and MCAM at distal ends, whether or not they were oriented toward CXCL12. In no case were cells oriented with their long axis perpendicular to the gradient, when Golgi and MCAM were distal to each other. Thus, Wnt5a induces polarization in dispersed cells, and cell polarization and directionality can be defined by the formation of W-RAMP structures and their localization relative to Golgi. Chemotaxis toward CXCL12 is not an underlying cause of Wnt5a-induced polarization, although polarization re-

quires interactions with external factors that confer cues for positional orientation.

Live-cell imaging showed that the W-RAMP structure is dynamic. Endogenous MCAM, labeled with antibody-coupled fluorophore, appeared polarized in cells within 25 min after Wnt5a addition. Enrichment of MCAM at one edge of the cell was in each case accompanied by rapid retraction of the membrane toward the cell body and nuclear movement in the same direction, revealing coordination between membrane retraction and directional cell movement (Fig. 2A and movies S1 to S4). Similar behavior was observed in cells transfected with MCAM fused to green fluorescent protein (MCAM-GFP) and actin fused to monomeric red fluorescent protein (mRFP-actin) (fig. S3 and movies S5 and S6).

The sequence of events in cells responding to Wnt5a was distinct from untreated cells. In a random sampling of live cells, each monitored for 100 min, the W-RAMP structure appeared in 36% of cells treated with Wnt5a (8 of 22 cells), with an average duration of 16 min. Wnt5a-treated cells often showed multiple waves, with the first wave appearing within 10 min, and the majority associated with membrane retraction (Fig. 2, B and C, and movies S7 to S9). In contrast, W-RAMP
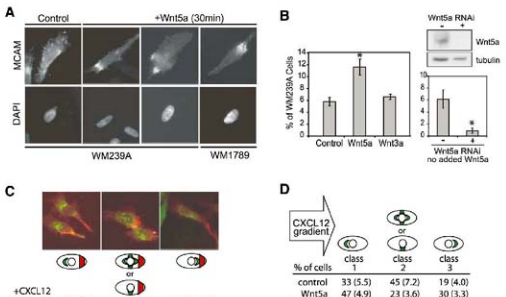
Fig. 1. An asymmetric "W-RAMP structure" forms in response to Wnt5a and defines cell polarization and orientation. (**A**) Examples of cells with uniform MCAM localization versus asymmetric W-RAMP structures, observed by indirect immunofluorescence in WM239A cells treated for 30 min with or without Wnt5a or in WM1789 cells treated for 30 min with Wnt5a. Nuclei are shown by 4′,6′-diamidino-2-phenylindole (DAPI) staining. (**B**) Wnt5a significantly increases the number of WM239A cells displaying asymmetric MCAM compared with control or Wnt3a-treated cells. [Data show means ± SEM. *P < 0.005 by Student's t test. Number of experiments (n) = 4. Cells counted: control, 1391; Wnt5a, 1380; Wnt3a, 1408.] RNAi knockdown of endogenous Wnt5a reduces the number of cells exhibiting the W-RAMP structure under untreated conditions, from 6% to less than 0.9% (Data show means with highest and lowest values, n = 2. Cells counted: control, 648; Wnt5a-RNAi, 1051). (**C**) In the presence of CXCL12, Wnt5a increased the percentage of cells with W-RAMP structures (red) positioned distal to Golgi (green) (means ± SEM; *P < 0.005, n = 4. Cells counted: control, 99; Wnt5a, 136). In the absence of CXCL12, Wnt5a did not affect the W-RAMP structure position relative to Golgi (n = 3. Cells counted: control, 134; Wnt5a, 131). (**D**) Wnt5a increased the number of cells with Golgi pointed toward or away from the gradient (class 1 and 3), with little directional bias.

structures were found in only 11% of untreated cells (3 of 27 cells), none of which formed multiple waves. Single waves appeared after 20 min with an average duration of 7 min and were not associated with membrane retraction (Fig. 2C, and movies S10 to S12). This revealed ~40% penetrance of the Wnt5a response within an asynchronous cell population and confirmed that, in untreated cells, the lower frequency quantified at 30 min (12 to 15%) can be ascribed to the transient properties of the W-RAMP structure.

In epithelial cells, tail-end recruitment of F-actin and myosin IIB triggers membrane retrac-

tion (12). We therefore examined the distribution of these proteins with respect to MCAM. In untreated melanoma cells, F-actin and myosin IIB were uniformly distributed throughout the cell. However, after Wnt5a treatment, F-actin formed a dense band near each W-RAMP structure, with close associations between F-actin and MCAM extending to the cell edge (Fig. 3A). Myosin IIB also localized to this region, overlapping MCAM and the dense band (Fig. 3A). Asymmetric F-actin and myosin IIB structures, similar to MCAM, were also observed in xenograft tumors (fig. S2B). Thus, the organization

of the W-RAMP structure involves close associations between cytoskeletal components and cell adhesion receptors. This parallels behavior in Drosophila wing epithelia, where actin and myosin II organize asymmetrically to regulate wing hair formation in response to Wnt signaling (13).

We examined the role of F-actin and myosin IIB in forming the W-RAMP structure. Cell treatment with latrunculin B at concentrations high enough to block dynamic actin reorganization, but low enough to preserve actin stress fibers, completely prevented Wnt5a-regulated
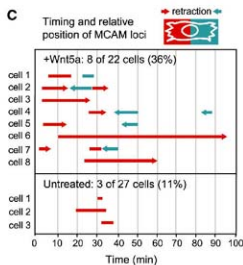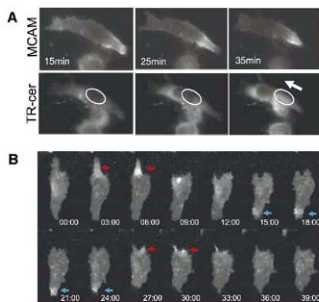
Fig. 2. Formation of the W-RAMP structure is dynamic and associated with membrane retraction. (**A**) Live WM239A cells were incubated with AlexaFluor-488–conjugated MCAM antibody and Texas Red (TR)–ceramide. In response to Wnt5a (in absence of CXCL12) MCAM decreases at the left end of the cell and increases at the right, followed by membrane retraction toward the W-RAMP structure. TR-ceramide staining reveals nuclear translocation in the direction of membrane retraction (arrow). (**B**) The W-RAMP structure forms in successive waves. MCAM-GFP expressed in WM239A cells accumulates at one cell end (red arrow) followed by membrane retraction (after 9 min). A second W-RAMP structure forms at the opposite cell end (blue arrow),

followed by membrane retraction (after 27 min). A third W-RAMP structure (red arrow) is followed by membrane retraction at 30 min. (**C**) Live cells were monitored continuously for MCAM-GFP localization. The x axis represents time after addition of Wnt5a. The length of each bar represents the duration of each wave of polarized accumulation of MCAM. Red versus blue lines distinguish the relative end of the cell at which the wave forms. Arrowheads indicate cases where membrane retraction follows formation of a wave, and the relative end of the cell at which the retraction occurs. Representative cells can be viewed in movies S7 to S12.
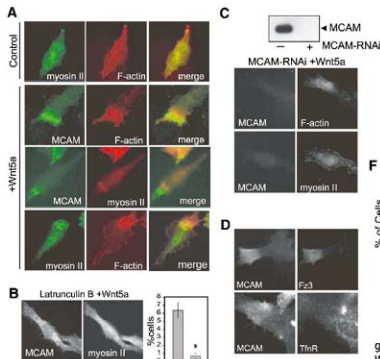
Fig. 3. The W-RAMP structure is formed through reciprocal interactions with actin and myosin IIB. (**A**) In Wnt5a-treated cells, myosin IIB and F-actin form asymmetric structures that partially overlap with the W-RAMP structure. (**B**) Latrunculin B (10 nM) pretreatment of cells strongly inhibits Wnt5a-dependent formation of the W-RAMP structure (means ± SEM; *P < 0.005. Cells counted: dimethylsulfoxide-treated, 1362; LatB, 1538). (**C**) RNAi knockdown of MCAM expression leads to actin structures that are more diffuse and disorganized than controls, as well as complete abrogation of myosin IIB structures. (**D**) Indirect immunofluorescence of WM239A cells shows that Fz3 colocalizes with the W-RAMP structure, but transferrin receptor does not. (**E**) Pull-down assay with immobilized

glutathione S-transferase–rhotekin reveals elevated RhoB-GTP in response to Wnt5a. Dvl2 knockdown blocks the activation of RhoB by Wnt5a. (**F**) Dvl2 RNAi knockdown abolishes W-RAMP structure in Wnt5a-treated cells (*P < 0.05, n = 3. Cells counted: control RNAi, 1002; Dvl2-RNAi, 1139). (**G**) Transient transfection of cells with DN-RhoB blocks Wnt5a-dependent formation of the W-RAMP structure (*P < 0.05, n = 3. Cells counted: WT-RhoB, 608; DN-RhoB, 590).

asymmetry of MCAM and myosin IIB without affecting overall cell morphology (Fig. 3B). Conversely, blocking MCAM expression by treating cells with MCAM-RNAi before adding Wnt5a completely abolished the dense actin and myosin IIB bands in all cells, while allowing a more diffuse asymmetry of actin (Fig. 3C and fig. S4). This demonstrates that polarization of the W-RAMP structure requires actin polymerization and that reciprocal interactions between MCAM and actin mediate recruitment of myosin IIB. We hypothesize that the transient and dynamic movement of the W-RAMP structure triggers localized membrane retraction by directing recruitment of actin and myosin IIB. Membrane retraction subsequently promotes nuclear movement, establishing the directionality of cell locomotion.

Fz3, a noncanonical Wnt receptor, was also recruited to the W-RAMP structure, but transferrin receptor, which is not involved in Wnt signaling, was not (Fig. 3D). Depletion of disheveled-2 (Dvl2) by RNAi and inhibition of PKC completely blocked the W-RAMP structure (Fig. 4, E and F, and fig. S5A). Thus, formation of the W-RAMP structure requires known effectors of noncanonical Wnt signaling.

We also examined Rho guanosine triphosphatases (GTPases), because RhoA, in many instances, is important for Wnt signaling (14). In vitro pull-down assays showed that Wnt5a enhanced the abundance of guanosine triphosphate (GTP)–bound active RhoB within 1 hour of stimulation, but did not affect the amount of RhoA-GTP (Fig. 3E and fig. S5, B and C). Expression of dominant-negative mutant (DN) RhoB(T19N) (in which Thr[19] is replaced by Asn) in WM239A cells followed by Wnt5a treatment suppressed the number of cells forming W-RAMP structures, down to amounts in untreated cells (Fig. 3G). Furthermore, inhibition of Dvl2 by RNAi blocked activation of RhoB, which confirmed that RhoB is a downstream effector of Wnt (Fig. 3E). These findings reveal the requirement for RhoB in Wnt5a signaling, for controlling the polarized distribution of MCAM and associated cytoskeletal proteins.

Because RhoB directs endosome trafficking, we asked whether MCAM is localized to intracellular membrane compartments and whether W-RAMP structures are regulated by Wnt5a-responsive vesicle trafficking (15). Wnt5a promotes membrane internalization (16), and such

behavior would parallel events in zebrafish gastrulation, in which directional mesendodermal cell movement in response to Wnt11 requires endocytosis of E-cadherin and can be blocked by GTPase-deficient mutant dynamin or DN Rab5 (17). Confocal imaging showed that polarized MCAM and F-actin structures were excluded from nuclei, but were otherwise present throughout the cell body and not restricted to the cell membrane (Fig. 4A). Visualizing W-RAMP–containing cells by electron microscopy revealed that the cell region that included the W-RAMP structure was dense with multivesicular bodies (MVBs) decorated with filamentous structures (Fig. 4B), which suggested that the W-RAMP structure is composed of MVBs associated with MCAM and polymerized cytoskeletal proteins. These findings are consistent with the requirement of the W-RAMP structure for RhoB, which is known to localize to and direct trafficking of MVBs (15).

We tested whether formation of the W-RAMP structure depended on receptor internalization and endosomal trafficking. A DN of dynamin (K44A, in which Lys[44] is replaced by Ala) blocks receptor-mediated endocytosis by inhibiting plasma membrane cleavage at the neck of invaginat-
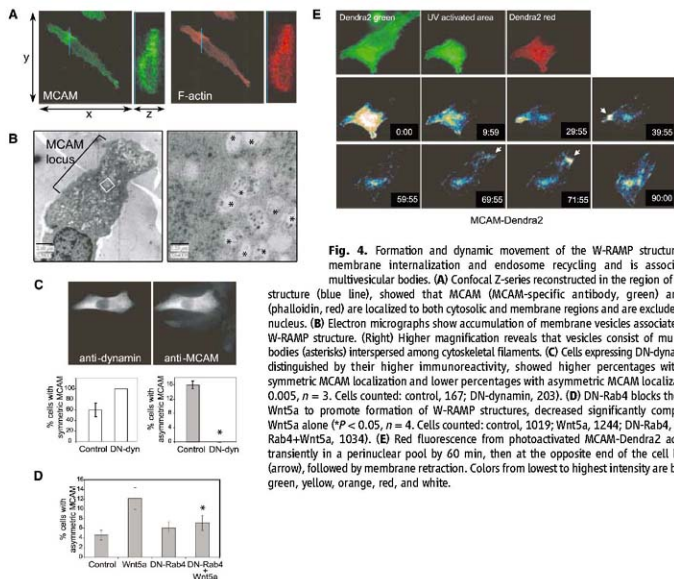
Fig. 4. Formation and dynamic movement of the W-RAMP structure requires membrane internalization and endosome recycling and is associated with multivesicular bodies. (A) Confocal Z-series reconstructed in the region of the MCAM structure (blue line), showed that MCAM (MCAM-specific antibody, green) and F-actin (phalloidin, red) are localized to both cytosolic and membrane regions and are excluded from the nucleus. (B) Electron micrographs show accumulation of membrane vesicles associated with the W-RAMP structure. (Right) Higher magnification reveals that vesicles consist of multivesicular bodies (asterisks) interspersed among cytoskeletal filaments. (C) Cells expressing DN-dynamin, easily distinguished by their higher immunoreactivity, showed higher percentages with uniform symmetric MCAM localization and lower percentages with asymmetric MCAM localization (*P < 0.005, n = 3. Cells counted: control, 167; DN-dynamin, 203). (D) DN-Rab4 blocks the ability of Wnt5a to promote formation of W-RAMP structures, decreased significantly compared with Wnt5a alone (*P < 0.05, n = 4. Cells counted: control, 1019; Wnt5a, 1244; DN-Rab4, 1262; DN-Rab4+Wnt5a, 1034). (E) Red fluorescence from photoactivated MCAM-Dendra2 accumulates transiently in a perinuclear pool by 60 min, then at the opposite end of the cell by 71 min (arrow), followed by membrane retraction. Colors from lowest to highest intensity are black, blue, green, yellow, orange, red, and white.

ing vesicles (18). Cells transfected with DN dynamin completely suppressed MCAM asymmetry, which resulted in uniform distribution of MCAM across the cell (Fig. 4C). Thus, membrane internalization, an initial step in endosomal trafficking, was necessary to form the W-RAMP structure.

Because Rab4 directs early endosomes to the recycling endosome compartment, cells were transfected with DN-Rab4 (N121I) (in which Asn[121] is replaced by Ile) fused to GFP in order to block endosome trafficking. DN-Rab4 blocked formation of W-RAMP structures down to control levels (Fig. 4D), which indicated that Rab4 mediates Wnt5a signaling, as observed with Dvl2, dynamin, and RhoB. Although GFP-Rab4 was not observed within W-RAMP structures, MCAM partially overlapped GFP-Rab4 within the perinuclear region, consistent with its association with recycling endosomes (fig. S6). The results indicate that dynamic movement and intracellular translocation of MCAM is mediated via internalization of MCAM and trafficking of receptor endosomes. This might occur through a linear pathway with Wnt5a upstream of Rab4 and other effectors, or through parallel pathways involving convergence between Wnt5a and endosomal trafficking effectors.

Cells forming W-RAMP structures typically displayed only one structure in each cell at any moment. This suggested that dynamic redistribution of W-RAMP structures may involve release of MCAM from one region of the cell, movement to distal regions, and assembly into another structure. To test this, cells were transfected with MCAM fused to Dendra2, a photoactivatable protein, which, upon ultraviolet (UV) illumination, switches from green to red fluorescence (19). One-third of the cell area was UV-illuminated, followed by Wnt5a addition (Fig. 4E

and movie S13). Within 15 to 30 min, photoactivated MCAM-Dendra2 decreased within the illuminated region and, over 30 to 60 min, accumulated within a perinuclear compartment consistent with the location of Rab4-positive recycling endosomes. By 60 min, MCAM-Dendra2 appeared in punctate cytosolic patterns within the non–UV-illuminated region, followed by its accumulation at the cell edge and subsequent membrane retraction. This confirms that the W-RAMP structure is subject to turnover, where proteins are recycled in a process of disassembly, intracellular redistribution, and reassembly. Given the requirement for RhoB, dynamin, and Rab4, as well as the association of MVBs with the W-RAMP structure, we speculate that turnover involves movement of MCAM through recycling endosomes and MVBs and intracellular translocation of MVBs.

In conclusion, we report a mechanism by which dispersed cells respond acutely to noncanonical Wnt signaling; it involves recruitment and redistribution of cellular proteins into an intracellular structure that integrates receptors for cell adhesion and cell signaling with components of the cytoskeletal architecture. In the presence of gradient cues from secreted factors, the W-RAMP structure asymmetrically distributes in a polarized manner, where it directs membrane retraction and thus influences the direction of cell movement. This allows Wnt5a to control polarity and directional orientation, even in cells lacking positional information from cell-cell contacts. The W-RAMP structure requires Dvl2 and PKC, involves membrane internalization and endosome trafficking and is regulated by RhoB and formation of MVBs. This contrasts with other mechanisms in which Wnt regulates cytoskeletal

architecture via RhoA and receptor distribution via endocytic pathways. Our findings add insight to the understanding of acute intracellular events mediated by Wnt signaling.

References and Notes
1. P. N. Adler, Dev. Cell 2, 525 (2002).
2. Y. Shimada, S. Yonemura, H. Ohkura, D. Strutt, T. Uemura, Dev. Cell 10, 209 (2006).
3. F. D. Park, J. R. Tenlen, J. R. Priess, Curr. Biol. 14, 2252 (2004).
4. B. Goldstein, H. Takeshita, K. Mizumoto, H. Sawa, Dev. Cell 10, 391 (2006).
5. M. A. Hilliard, C. I. Bargmann, Dev. Cell 10, 379 (2006).
6. J. Whangbo, C. Kenyon, Mol. Cell 4, 851 (1999).
7. A. T. Weeraratna et al., Cancer Cell 1, 279 (2002).
8. Materials and methods are available as supporting material on Science Online.
9. K. Satyamoorthy, J. Muyrers, F. Meier, D. Patel, M. Herlyn, Oncogene 20, 4676 (2001).
10. R. A. Bartolome et al., Cancer Res. 64, 2534 (2004).
11. T. Murakami et al., Cancer Res. 62, 7328 (2002).
12. J. Kalega, Mol. Biol. Cell 14, 4745 (2003).
13. C. G. Winter et al., Cell 105, 81 (2001).
14. M. T. Veeman, J. D. Axelrod, R. T. Moon, Dev. Cell 5, 367 (2003).
15. S. Ellis, H. Mellor, Trends Cell Biol. 10, 85 (2000).
16. W. Chen et al., Science 301, 1391 (2003).
17. F. Ulrich et al., Dev. Cell 9, 555 (2005).
18. A. M. Van der Bliek et al., J. Cell Biol. 122, 553 (1993).
19. N. G. Gurskaya et al., Nat. Biotechnol. 24, 461 (2006).
20. We are indebted to D. Chan and Z. Zhang for help with xenograft tumor growth, N. Camp for preliminary data collection, and T. Giddings for guidance with TEM. This work was supported by NIH grants F32-CA112847 (E.S.W.), F32-CA105796 (G.M.A.), and R01-CA118972 (N.G.A.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5874/365/DC1
Materials and Methods
Figs. S1 to S6
References and Notes
Movies S1 to S13

2 October 2007; accepted 5 March 2008
10.1126/science.1151250

A Model for Neuronal Competition During Development

Christopher D. Deppmann,[1,2]* Stefan Mihalas,[1,3]* Nikhil Sharma,[1,2]*
Bonnie E. Lonze,[1,2] Ernst Niebur,[1,3] David D. Ginty[1,2]†

We report that developmental competition between sympathetic neurons for survival is critically dependent on a sensitization process initiated by target innervation and mediated by a series of feedback loops. Target-derived nerve growth factor (NGF) promoted expression of its own receptor TrkA in mouse and rat neurons and prolonged TrkA-mediated signals. NGF also controlled expression of brain-derived neurotrophic factor and neurotrophin-4, which, through the receptor p75, can kill neighboring neurons with low retrograde NGF-TrkA signaling whereas neurons with high NGF-TrkA signaling are protected. Perturbation of any of these feedback loops disrupts the dynamics of competition. We suggest that three target-initiated events are essential for rapid and robust competition between neurons: sensitization, paracrine apoptotic signaling, and protection from such effects.

In the developing peripheral nervous system, target tissues specify the amount of innervation they receive by secreting limiting amounts of neurotrophic factors, which are required for neuronal survival (1, 2). The prototypic target

cell–derived neurotrophic factor, NGF, supports survival of developing sympathetic and cutaneous sensory neurons (3, 4). NGF engages its receptor tyrosine kinase TrkA on the distal axon, and the NGF-TrkA complexes travel retrogradely from

the periphery to the cell body, where they induce prosurvival signaling programs and NGF-dependent transcriptional programs (5–10).

To identify NGF-dependent transcriptional events that may enable one neuron to gain competitive advantage over its neighbor with similar potential, we performed a comprehensive in vivo comparison of gene expression profiles of sympathetic neurons isolated from newly born mice with an intact or ablated NGF locus (11). To circumvent the requirement of NGF for survival, we used Bax[/−] animals, which are deficient in sympathetic neuron apoptosis during development (12–14). Expression of brain-derived neurotrophic factor (BDNF), TrkA, and another

[1]The Solomon Snyder Department of Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [2]Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [3]The Zanvyl Krieger Mind/Brain Institute, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: dginty@jhmi.edu

neurotrophin receptor, p75, was decreased in the absence of NGF (table S1 and fig. S1A). NGF dependence of TrkA expression was confirmed by in situ hybridization, immunohistochemistry, reverse transcriptase polymerase chain reaction (RT-PCR), and immunoblot analysis (Fig. 1, A and B, and fig. S1B). NGF also regulated expression of krupple-like factor 7, a transcription factor that regulates TrkA expression (*15*) (Fig. 1B). Moreover, TrkA expression and downstream signaling were greatly reduced when cultured sympathetic neurons were deprived of NGF for 24 or 48 hours (Fig. 1C). We also compared the abundance of TrkA, activated phosphoTrkA (p-TrkA), and p-Akt in neurons with different maturities and found that the more mature neurons exhibited dramatic increases in all three markers of NGF prosurvival signaling (fig. S1C). In addition to the difference in NGF signaling magnitude, the duration of prosurvival signaling increased as a function of NGF exposure and neuron maturity (Fig. 1, C and D). Prosurvival signaling remained robust for at least 60 min after treatment with NGF in neurons deprived of NGF for 12 hours before treatment; neurons deprived of NGF for longer periods of time lost their ability to sustain prosurvival signaling. Taken together, these observations suggest the possibility that NGF induces feedback loops, in part through regulation of TrkA expression, that control magnitude and duration of NGF signaling.

To determine whether NGF-dependent gene expression events and resultant feedback loops underlie the acquisition of competitive advantage, we turned to a computational approach to model this process. Our model assigns one differential equation per neuron, representing the relative magnitude of trophic signaling, which we define as the amount of NGF-bound TrkA within the neuron. Another differential equation is assigned for the concentration of NGF available at the target, with the assumption that the rate of NGF production is constant. The outputs of these equations changed as a function of time (figs. S2 and S3). Computer simulations were used to relate these two equations to neuronal survival; a neuron was considered dead if it reached a trophic signaling state of 10% or less of the maximum value. The results of these simulations are represented in several forms, all with respect to time: (i) trophic signaling for each modeled neuron (Fig. 2, A to C), (ii) average trophic signaling (Fig. 2D), (iii) total neuron number (Fig. 2E), and (iv) NGF concentration at the target (Fig. 2F).

To assess the importance of NGF-dependent changes in both TrkA expression and signal duration during neuronal competition, we conducted a simulation in which these terms were fixed and independent of exposure to NGF. In this paradigm, all neurons rapidly reached a trophic steady state, and no competition occurred, which resulted in the survival of all neurons (Fig. 2, A and E). If either TrkA expression or signal duration was fixed and the other was allowed to

change in relation to NGF exposure, competition also failed to occur (Fig. 2, B and E, and fig. S4). Both signal strength and duration had to change after neuronal exposure to NGF in order for effective competition to occur (Fig. 2, C and E). When competition occurred, some neurons obtained a high trophic signaling state and others reached a low trophic signaling state and ultimately died (Fig. 2E and fig. S4K). In this system, survival outcomes were fairly independent of starting conditions (figs. S5 and S6). However, changes in NGF production, NGF-dependent TrkA expression, and TrkA signal duration had large effects on the dynamics of neuron elimination (figs. S7 to S9). Taken together, these modeling data indicate that target innervation–dependent changes in TrkA expression and signal duration may be essential for neuronal competition.

Clues from a model for synapse competition at the neuromuscular junction (*16*), as well as findings from our microarray analysis (fig. S1A

and table S1), led us to test whether, during competition for survival, neurons with high amounts of trophic signaling actively promote apoptosis of neurons with low trophic signaling. The putative apoptotic cue should have several characteristics: (i) It should kill neurons with low trophic signaling, and (ii) it should emanate from neurons with strong trophic signaling. (iii) Neurons with strong trophic signaling should be impervious to the apoptotic cue, and (iv) apoptotic cue signaling should commence after expression of TrkA increased. BDNF and neurotrophin-4 (NT4) satisfy the criteria for the apoptotic cues: (i) BDNF and NT4 can promote apoptosis of sympathetic neurons through the receptor p75 (*17*, *18*) (Fig. 3, C and D), and (ii) their expression is regulated by NGF in sympathetic neurons (Fig. 3, A and B, and table S1). (iii) Strong NGF-TrkA signaling blocks p75-mediated killing of sympathetic neurons (*19*) (Fig. 3F). (iv) Finally, p75 expression in sympathetic neurons commences in vivo 2 days after TrkA expression (*20*), and this expression is
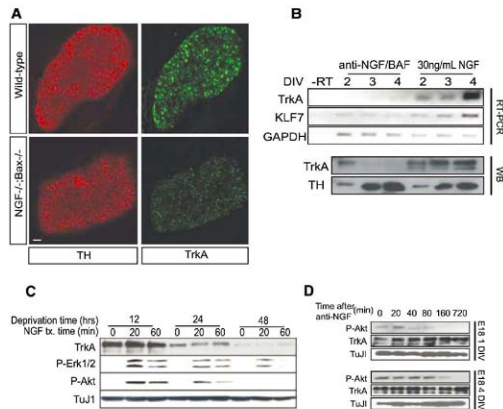
Fig. 1. NGF promotes TrkA expression and the duration of TrkA signaling events. (**A**) Immunohistochemistry for TrkA and tyrosine hydroxylase (TH) on SCGs from wild-type and $NGF^{-/-};Bax^{-/-}$ animals. (**B**) TrkA, KLF7, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) levels were measured by RT-PCR and TrkA and TH by Western blot (WB) after treatment of cultured rat sympathetic neurons with NGF (30 ng/ml) for indicated times. (**C**) TrkA, p-ERK, p-Akt, and TuJl measured by WB in P0.5 rat sympathetic neurons that were deprived of NGF for the indicated times followed by reexposure to NGF (30 ng/ml) for 20 or 60 min. (**D**) Sympathetic neuron cultures from E18 rats were grown in the presence of 30 ng/ml NGF for 1 day in vitro (DIV) or 4 DIVs. NGF-specific antibody (anti-NGF) was added, and NGF signaling events were assessed at the indicated times after NGF deprivation. Graph compares relative p-Akt levels for this experiment. Scale bar in (A), 50 μm. *$P < 0.05$ analysis of variance (ANOVA) followed by Tukey's post hoc test.

Fig. 2. Requirement for both NGF-dependent TrkA expression and signal duration during competition revealed by computer simulations. Shown are model simulation results for 100 neurons under different competitive conditions. (**A** to **C**) Simulated trophic signaling state of 100 individual neurons with indicated parameters over time. (A) TrkA production and signaling duration were held constant. (B) TrkA production was NGF-dependent but signaling duration was held constant. (C) TrkA production and signaling duration were NGF-dependent. (**D** to **F**) Comparison of various dynamic elements as a function of time in simulations from (A) to (C). (D) Average trophic signal strength; (E) cell survival; (F) relative amount of NGF at the target. The black, green, and blue lines in (D), (E), and (F) represent results of simulations described for (A), (B), and (C).
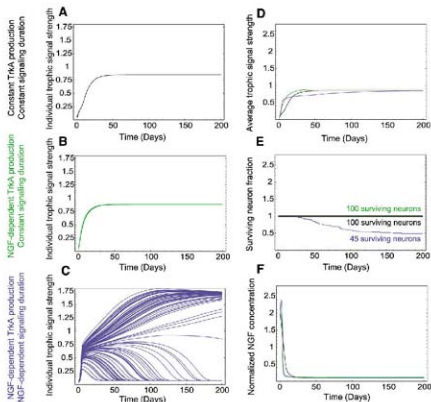
Fig. 3. Antagonism of retrograde NGF survival signaling by NGF induced BDNF and NT4 expression in sympathetic neurons. (**A**) RT-PCR for BDNF, NT-4, and GAPDH with RNA from NGF-treated (24 hours) or NGF-deprived sympathetic neurons from P0 mice grown 3 DIVs. (**B**) In situ hybridization for BDNF and NT-4 in wild-type and *NGF⁻/⁻;Bax⁻/⁻* animals at P0. Scale bar, 50 μm. (**C** to **F**) Neurotrophin promoted p75-dependent cell death of cultured sympathetic neurons. Indicated neurotrophins were applied for 36 hours, cell survival was determined by Hoechst staining, and results are means ± SEM (n = 4 experiments). (C) Survival of P0 to P2 rat sympathetic neurons maintained in medium containing NGF (1 ng/ml). Green dashed line represents maximum survival, and red dashed line indicates maximum death, both conditions were assessed after 36 hours with either 1 ng/ml of NGF or NGF-specific antibody (anti-NGF), respectively. (D) Survival of P0 to P2 *p75⁻/⁻* or wild-type mouse sympathetic neurons treated with NGF (1 ng/ml) or NGF-specific antibody (αNGF). (**E** and **F**) P0 sympathetic neurons grown in compartmentalized chambers (represented by illustration left of graphs) for 5 to 7 days before medium was changed to contain NGF (5 ng/ml) on the distal axons. The indicated neurotrophins or NGF-specific antibody (αNGF) was applied to the cell bodies for 36 hours before assessing survival. Unless otherwise indicated, BDNF and NT-4 were applied at a concentration of 250 ng/ml *P < 0.01 ANOVA followed by Tukey's post hoc test.
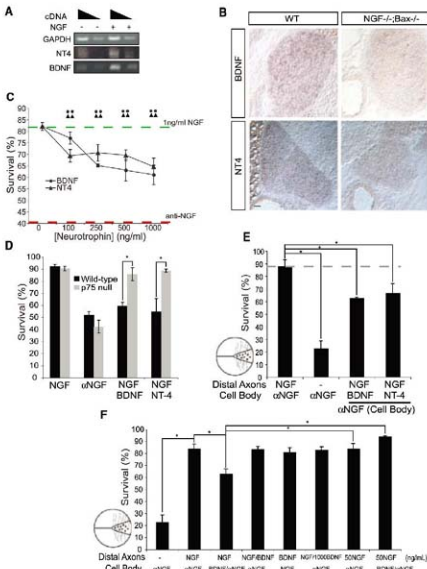
regulated by NGF (21, 22) (table S1). Thus, we propose that BDNF and NT4 are NGF-regulated apoptotic cues that developing sympathetic neurons and that innervation-dependent expression of p75 regulates a neuron's susceptibility to these signals.

We next sought to define other features of p75-mediated cell death. BDNF and NT4 caused dose-dependent killing of sympathetic neurons in the presence of concentrations of NGF insufficient to promote maximal activation of TrkA (Fig. 3C). In p75$^{-/-}$ neurons, these neurotrophins had no effect on survival (18) (Fig. 3D). Next, the spatial properties of these apoptotic signals were characterized in a compartmentalized culture

system (23). If a low concentration of NGF was applied exclusively to distal axons and either BDNF or NT4 was applied to cell bodies, NGF-dependent survival was reduced to a similar extent as that observed in mass cultures (Fig. 3, C and E). In contrast, if BDNF was applied exclusively to distal axons, cell death was not observed, which suggested that apoptotic signals must be produced and act at the cell body to affect cell survival (Fig. 3F). It is noteworthy that induction of cell death by BDNF-p75 signaling on the cell body was suppressed when larger amounts of NGF were added to distal axons (Fig. 3F).

To determine the potential benefits of NGF-dependent expression of signals that promote

death of neurons with low trophic signaling, we added several semiquantitative equations to the computational model presented in Fig. 2C to characterize the apoptotic signals and their receptors. As determined experimentally for BDNF and NT4, the variable representing the amount of apoptotic signal increases with neuronal exposure to NGF and has no effect on neurons with high retrograde NGF-TrkA signaling. Computer simulations revealed that the addition of this parameter hastened, by more than 10-fold, the killing of those neurons that did not gain competitive advantage through the sensitization process (Fig. 4, A and B). The extent to which the apoptotic signals hastened neuronal death was

Fig. 4. Mathematical modeling predicts changes in competition dynamics that are corroborated in vivo. (A and B) Simulations of neuronal competition with and without NGF-dependent production of an apoptotic cue (red and blue, respectively). (A) Trophic signaling strength of 100 individual neurons as a function of time. (B) Cell survival as a function of time. (C) Cell counts of Nissl-stained SCG sections from p75$^{-/-}$ or wild-type mice at indicated developmental ages. Results are means ± SEM (n = 3 for each age). (D) Immunostaining for TH and TrkA in SCGs from Bax$^{-/-}$ and wild-type animals. Scale bar, 50 μm. (Right) Boxed region in middle panels magnified 5×. (E and F) Quantification of soma size (E) and relative TrkA amount (F) in P5 Bax$^{-/-}$ versus wild-type SCGs. Results are represented as a percentage of total neurons counted. *P < 0.01 ANOVA followed by Tukey's post hoc test. (G) Model for developmental competition: 1) Before target innervation neurons are modestly responsive to NGF; 2) upon target innervation and exposure to NGF, levels of TrkA, then BDNF and NT-4 are increased; 3) induction of p75 expression, as well as differential sensitization of neurons, by modulation of NGF-TrkA signal strength and duration. 4) BDNF and NT-4 (apoptotic cues) kill neurons with low NGF-TrkA signaling; neurons with high NGF-TrkA signaling are resistant; 5) selection and neuronal death.
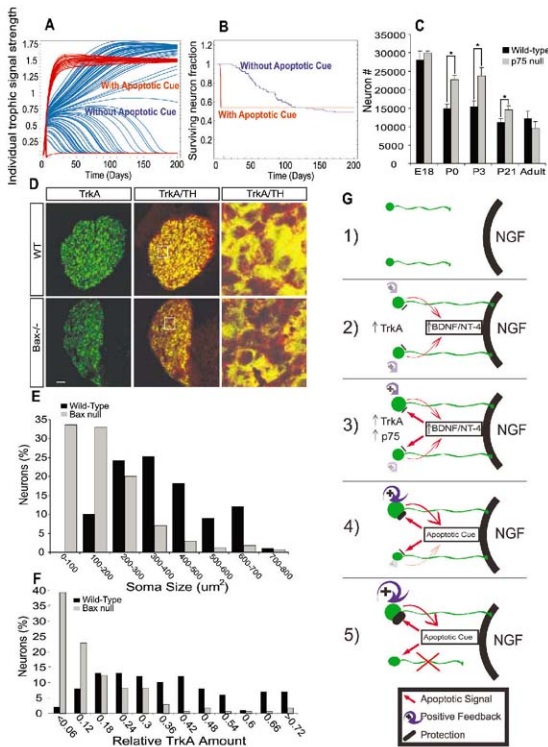
dependent on how much NGF-TrkA signaling was required for protection and the rate of production of BDNF and NT4 in response to NGF (figs. S10 and S11).

We thus propose a model (Fig. 4G) in which, before target innervation, sympathetic neurons are modestly responsive to NGF because of low TrkA levels. Upon target innervation, neurons acquire NGF, and subtle differences in initial amounts of NGF signaling are amplified through transcription-dependent feedback loops into large cell-autonomous differences in both strength and duration of TrkA signaling, which are the ultimate determinants for whether a neuron lives or dies. We further propose that expeditious competition requires target innervation–dependent expression of apoptotic cues (BDNF and NT4), susceptibility to such signals (expression of p75), and protection from apoptotic signals (strong retrograde NGF-TrkA signaling) (Fig. 4G and fig. S12). These features enable competition that is rapid, robust, and stable, even in a scenario in which all neurons arrive at their target simultaneously and are virtually equivalent in their initial responsiveness to target-derived NGF.

Several predictions arise from this model. One prediction is that competing neurons distinguish themselves from one another on the basis of their amount of TrkA expression; those with abundant TrkA live, whereas those with low amounts die. Consistent with this prediction, in $Bax^{-/-}$ mice at postnatal day 5 (P5), over half of the neurons had extremely low levels of TrkA and small soma areas, whereas the rest of these neurons had larger amounts of TrkA and cell body areas comparable to those of wild-type controls (Fig. 4, D to F, and figs. S13 and S14). The model also predicts that atrophic neurons in superior cervical ganglia (SCG) from P5 $Bax^{-/-}$ mice should display persistent p75-mediated apoptotic signals given that these are neurons that would normally have died during developmental competition. Indeed, compared with neurons with high TrkA expression, atrophic neurons in $Bax^{-/-}$ ganglia have large amounts of p75 and p75-dependent signaling as measured by immunostaining of the phosphorylated c-Jun transcription factor (figs. S17 and S18).

Finally, the notion that NGF-dependent apoptotic signal hastens but does not change the fundamental elements of competition was tested in mice lacking p75. A developmental time course comparing numbers of SCG neurons in wild-type and $p75^{-/-}$ animals was performed at embryonic day 18 (E18), P0, P3, P21, and 6 months of age. As predicted by the modeling data (Fig. 4, A and B), SCGs from wild-type and $p75^{-/-}$ animals had a similar number of neurons at E18, whereas P0 and P3 $p75^{-/-}$ SCGs had ~40% excess of neurons compared with wild-type animals. At 6 months of age, $p75^{-/-}$ SCGs had the same number of neurons as control SCGs (Fig. 4C) [see also (18, 24)].

These genetic and computational modeling data show that two neurotrophin receptors and potentially three neurotrophin ligands orchestrate developmental competition between SCG neurons through a series of feedback loops. Although the cast of players for other populations of competing neurons is almost certain to be different from those that coordinate competition in the SCG, we suggest that the underlying principles of target initiated sensitization, paracrine apoptotic signaling, and protection from apoptotic signals are common features of neuronal competition at large.

References and Notes
1. V. Hamburger, R. Levi-Montalcini, *J. Exp. Zool.* **111**, 457 (1949).
2. R. Levi-Montalcini, *Science* **237**, 1154 (1987).
3. R. Levi-Montalcini, B. Booker, *Proc. Natl. Acad. Sci. U.S.A.* **46**, 384 (1960).
4. C. Crowley *et al.*, *Cell* **76**, 1001 (1994).
5. H. M. Heerssen, M. F. Pazyra, R. A. Segal, *Nat. Neurosci.* **7**, 596 (2004).
6. A. Riccio, S. Ahn, C. M. Davenport, J. A. Blendy, D. D. Ginty, *Science* **286**, 2358 (1999).
7. F. L. Watson *et al.*, *Nat. Neurosci.* **4**, 981 (2001).
8. H. Ye, R. Kuruvilla, L. S. Zweifel, D. D. Ginty, *Neuron* **39**, 57 (2003).
9. J. D. Delcroix *et al.*, *Neuron* **39**, 69 (2003).
10. C. L. Howe, W. C. Mobley, *Curr. Opin. Neurobiol.* **15**, 40 (2005).
11. Materials and methods are available as supporting material on *Science* Online.
12. N. O. Glebova, D. D. Ginty, *J. Neurosci.* **24**, 743 (2004).
13. T. D. Patel, A. Jackman, F. L. Rice, J. Kucera, W. D. Snider, *Neuron* **25**, 345 (2000).
14. T. L. Deckwerth *et al.*, *Neuron* **17**, 401 (1996).
15. L. Lei, L. Ma, S. Nef, T. Thai, L. F. Parada, *Development* **128**, 1147 (2001).
16. J. R. Sanes, J. W. Lichtman, *Annu. Rev. Neurosci.* **22**, 389 (1999).
17. E. C. Yeiser, N. J. Rutkoski, A. Naito, J. Inoue, B. D. Carter, *J. Neurosci.* **24**, 10521 (2004).
18. S. X. Bamji *et al.*, *J. Cell Biol.* **140**, 911 (1998).
19. M. Majdan, G. S. Walsh, R. Aloyz, F. D. Miller, *J. Cell Biol.* **155**, 1275 (2001).
20. S. Wyatt, A. M. Davies, *J. Cell Biol.* **130**, 1435 (1995).
21. R. Kuruvilla *et al.*, *Cell* **118**, 243 (2004).
22. F. D. Miller, T. C. Mathew, J. G. Toma, *J. Cell Biol.* **112**, 303 (1991).
23. R. B. Campenot, *Methods Enzymol.* **58**, 302 (1979).
24. C. Brennan, K. Rivas-Plata, S. C. Landis, *Nat. Neurosci.* **2**, 699 (1999).
25. We thank C. Jie and F. Martinez Murillo of the Johns Hopkins Medical Institute microarray core for their assistance in microarray analysis; and N. Galano, A. Harrington, R. Kuruvilla, A. Levchenko, W. Luo, and K. Wright for helpful comments during manuscript preparation. This work is supported by NIH fellowship NS053187 (C.D.D.), A Woodrow Wilson Undergraduate Research Fellowship (N.S.), and NIH grants NS34814 (D.D.G.) and EY016281 (E.N.). D.D.G. is an investigator of the Howard Hughes Medical Institute. The microarray analysis data have been deposited in National Center for Biotechnology Information's Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/), accession no. GSE10498.

Recapitulation of IVIG Anti-Inflammatory Activity with a Recombinant IgG Fc

Robert M. Anthony,[1] Falk Nimmerjahn,[1,4] David J. Ashline,[2] Vernon N. Reinhold,[2] James C. Paulson,[3] Jeffrey V. Ravetch[1]*

It is well established that high doses of monomeric immunoglobulin G (IgG) purified from pooled human plasma [intravenous immunoglobulin (IVIG)] confer anti-inflammatory activity in a variety of autoimmune settings. However, exactly how these effects are mediated is not clear because of the heterogeneity of IgG. Recent studies have demonstrated that the anti-inflammatory activity of IgG is completely dependent on sialylation of the N-linked glycan of the IgG Fc fragment. Here we determine the precise glycan requirements for this anti-inflammatory activity, allowing us to engineer an appropriate IgG1 Fc fragment, and thus generate a fully recombinant, sialylated IgG1 Fc with greatly enhanced potency. This therapeutic molecule precisely defines the biologically active component of IVIG and helps guide development of an IVIG replacement with improved activity and availability.

Although originally used as an antibody-replacement therapy, when given at high doses (1 to 2 g/kg), intravenous immunoglobulin (IVIG) has general anti-inflammatory properties and has been widely used to treat autoimmune diseases, including immune thrombocytopenia (ITP), rheumatoid arthritis, and systemic lupus erythematosus. The anti-inflammatory activity of IVIG has been demonstrated in a variety of animal models of autoimmunity, including autoantibody ITP (1), serum-transfer arthritis (2), and nephrotoxic nephritis (3) and is a property of the Fc fragment and its associated glycan (1, 4, 5). Removal of the terminal sialic acid from IVIG or its papain-derived Fc fragment abrogates the anti-inflammatory activity in these animal models. Conversely, enrichment of the sialylated fraction of IVIG enhances this activity (4).

Because sialic acid can be found in either a 2,3- or 2,6-linkage to the penultimate galactose on the complex, biantennary glycan found at Asn^{297} in immunoglobulin G (IgG) (Fig. 1A) (6), we set out to determine which linkage type was present in IVIG by analyzing the enzymatically released N-linked glycan from Fc fragments derived from IVIG (Fig. 1, B to D) (7). Sequential mass spectrometric–based glycan analysis revealed a preferential β-linkage in the IVIG preparations that conferred anti-inflammatory activity (Fig. 1, B to D). To confirm this association, we next analyzed FcγR activity after treatment with either 2,3- or 2,3/2,6-sialidase (SA) in the K/BxN serum-transfer arthritis model (8). Consistent with the glycan analysis, we observed complete abrogation of the anti-inflammatory activity when the

2,6-linkages were removed (Fig. 1E and fig. S1). In contrast, no 2,3-linkages were detected on intact IVIG, and the anti-inflammatory activity was unaffected after 2,3-SA treatment. These results are in agreement with previous studies that detected a preferential utilization of the 2,6-linkage of sialic acid–galactose in human IgG preparations and preferential expression of the 2,6-sialyltransferase (2,6ST) in B cells (9–13).

To determine whether the preferential 2,6-sialic acid–galactose linkage was a necessary and sufficient property of the anti-inflammatory activity of the sialylated, N-linked glycan at Asn^{297}, we generated IVIG-derived Fc fragments terminating in either the 2,3- or 2,6-sialic acid linkage. To efficiently sialylate the Fc fragment derived from IVIG, we first converted the population of glycans to the G2 (digalactosylated) (14, 15) form by in vitro treatment with β1,4-galactosyltransferase, resulting in a twofold increase in terminal galactose as determined by *Erythrina cristagalli* lectin (ECL) binding (Fig. 2, A and B). This galactosylated substrate was then reacted with either 2,3-sialyltransferase (2,3ST) or 2,6ST (Fig. 2A and fig. S2A). In vitro sialylation was evaluated directly by lectin blotting for 2,3- and 2,6-linkages (Fig. 2C) and indirectly by assaying for reduced

terminal galactose (ECL binding) (fig. S2, A and B). The 2,6ST quantitatively converted the G2 glycan to a fully sialylated form, as demonstrated by the absence of ECL reactivity (fig. S2). The efficiency of the 2,3ST was estimated to be ~50%, on the basis of residual ECL binding to these reacted Fc-linked glycans (fig. S2). ECL also bound the *Sambucus nigra* lectin (SNA)–enriched Fc derived from IVIG (SNA+ Fc), which contained ~30% fully sialylated Fc glycans (fig. S2), representing a 10-fold enrichment in sialic acid, as compared with unfractionated IVIG-derived Fc fragments. This level of sialylation observed in the SNA-enriched Fc preparation was sufficient to induce an anti-inflammatory response at a 10-fold reduced dose (as compared with unfractionated IVIG) and was demonstrated to confer enhanced anti-inflammatory activity through a pathway that requires the up-regulation of the inhibitory FcγR on effector inflammatory macrophages (3, 4).

We then administered these preparations to mice (table S1) in the K/BxN serum-induced arthritis model. Joint inflammation was effectively reduced by administering either the SNA+ IVIG Fc fragments, as has been observed previously (4), or the in vitro 2,6ST-treated IVIG Fc fragments

[1]Laboratory of Molecular Genetics and Immunology, The Rockefeller University, New York, NY 10021, USA. [2]Glycomics Center, Department of Biochemistry and Molecular Biology, University of New Hampshire, Durham, NH 03824, USA. [3]Departments of Chemical Physiology and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [4]Laboratory of Experimental Immunology and Immunotherapy, University of Erlangen-Nuremberg, Erlangen 91054, Germany.

*To whom correspondence should be addressed. E-mail: ravetch@rockefeller.edu
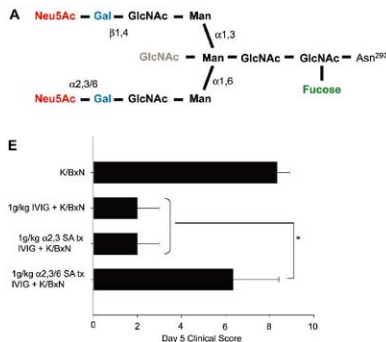
Fig. 1. α2,6-linkages are the predominant sialic acid linkages on IVIG Fc glycans. (**A**) The IgG Fc glycan is a bisecting core of seven sugars (black) and can vary at a number of positions by the addition of fucose (green) to the core, a bisecting GlcNAc (gray), or by the addition of galactose (blue) and sometimes sialic acid (red) to the arms. Neu5Ac, N-acetylneuraminic acid (sialic acid); Man, mannose. (**B**) Sequential mass spectrometry (MS) analysis of SNA-enriched IVIG Fc glycans was performed to determine the sialic acid linkage type and relative proportion of the linkages in the active component of IVIG. The resulting footprint of the B/Y galactosyl fragment monomer derived from the SNA-enriched Fc glycan was compared to the analogous B/Y fragments from (**C**) 2,3-sialyllactose and (**D**) 2,6-sialyllactose standards; the B/Y fragments derived from standards are depicted with the corresponding histogram. The spectrum generated from the SNA+ IVIG Fc glycan (B) most closely matches that of 2,6-sialyllactose (D), particularly with respect to the
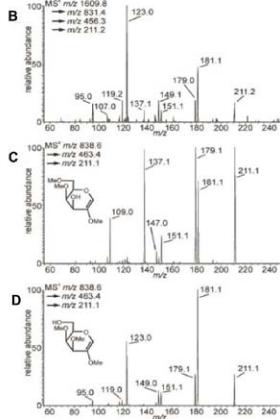
mass/charge ratio (*m/z*) 123 and 95 fragments and the much smaller abundances of *m/z* 137 and 109 ions. Next, IVIG was treated with linkage-specific SAs to remove only 2,3- (α2,3SA) or both 2,3- and 2,6- (α2,3/6SA) sialic acids. Me, methyl. (**E**) The SA-treated IVIG preparations were administered to mice prior to K/BxN sera, and footpad swelling was monitored over the next 7 days and recorded as clinical scores. Means and SDs (error bars) of five mice per group 5 days post-treatment are plotted. *$P < 0.05$, as determined by an analysis of variance (ANOVA) test, followed by a Tukey post hoc test.

(Fig. 3A). Suppression of inflammation was comparable with these two preparations and was 10-fold more active than the total Fc fragments derived from IVIG. In contrast, IVIG Fc fragments sialylated in vitro with 2,3ST were ineffective at reducing the joint inflammation induced by K/BxN serum (Fig. 3A), even at fourfold higher doses (fig. S2).

In contrast to the sialic acid–galactose linkage specificity for the Fc anti-inflammatory activity, the effect of sialylation of Fc receptor (FcR) binding and the consequent cytotoxicity was linkage-independent. Both terminal 2,3- and 2,6- in vitro sialylated IgG Fc reduced the cytotoxicity of an antiplatelet antibody, 6A6-IgG2b, in an in vivo model of ITP (Fig. 3B), consistent with previous studies (4, 16). Thus, the effect of Fc sialylation on the cytotoxicity of an IgG antibody is not dependent on the specificity of the linkage to the penultimate galactose. In contrast, the anti-inflammatory activity of the sialylated IgG Fc fragment [a property we have previously demonstrated to be independent of the canonical IgG Fc receptors (4, 5)] displayed a clear preference for the 2,6-sialic acid–galactose linkage (Fig. 3A and fig. S2B). These results further support our

previous observations that the anti-inflammatory property of IVIG is mediated through a distinct pathway (4, 5) that does not involve binding to canonical FcγRs.

To fully demonstrate that the in vivo anti-inflammatory activity of the 2,6-sialylated IgG Fc is solely a property of the IgG Fc glycan and not the result of other components that might be found in the heterogeneous IVIG Fc preparations, we set out to recapitulate the anti-inflammatory activity of sialylated IVIG Fc using a homogeneous, recombinant human IgG1 Fc substrate (rFc) derived from a cDNA expressed in 293T cells. The purified recombinant human IgG1 Fc fragment was glycan engineered in vitro (as described above) by β1,4-galactosylation, followed by 2,6-sialylation (Fig. 4A). The preparation was purified and characterized by lectin blotting (Fig. 4A) before in vivo analysis. As shown in Fig. 4B, the 2,6-sialylated recombinant human IgG1 Fc fragment demonstrated comparable anti-inflammatory activity to that obtained with either IVIG-derived sialic-enriched Fc fragments (SNA⁺ IVIG Fc) or in vitro 2,6-sialylated IVIG-derived Fc fragments. Each of these preparations was active at 30 mg/kg, as compared with

the 1000 to 2000 mg/kg required for native IVIG (table S1).

The exquisite specificity of glycan structures has long been appreciated as providing the structural basis for discrete biological responses, including lectins purified from plants (17), viruses (18–20), bacteria (18, 21), and parasites (22, 23). The results reported here demonstrate that the sialic acid–galactose linkage specificity for the N-linked glycan of the IgG Fc confers anti-inflammatory activity on the IgG and is a property of both the glycan and amino acid sequence (4).

The observation that the anti-inflammatory activity of IVIG is dependent on a precise glycan structure on the Fc further supports the model we have previously advanced (4, 5)—one in which a specific receptor for the sialylated Fc, and not a canonical Fc receptor, is involved in this pathway. Our data support a model in which the binding of the 2,6-sialylated Fc to its cognate receptor expressed on a population of splenic macrophages in the mouse results in the trans up-regulation of the inhibitory IgG Fc receptor on effector macrophages, located at sites of inflammation (such as the inflamed joint or glomerulus), thus raising the threshold required for
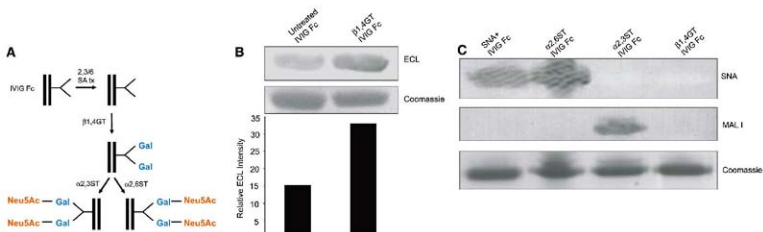
Fig. 2. In vitro sialylation of IVIG Fcs. (**A**) As shown in the schematic diagram of the sialylation strategy, IVIG Fc fragments were initially treated with 2,3/6 SA (to remove all sialic acid residues), galactosylated (with β1,4GT), and finally sialylated with either 2,3- or 2,6ST (α2,3ST and α2,6ST, respectively). (**B**) Galactosylation was verified by lectin blotting with ECL (top), which recognizes terminal galactose residues. Relative band intensity ratios of ECL to coomassie loading controls are plotted (bottom). (**C**) Galactosylated Fcs were then sialylated with 2,3- or 2,6ST, and each sialylation reaction was confirmed by lectin blotting for 2,6-linkages with SNA (top) and 2,3-linkages with *Maackia amurensis* lectin I (MAL I) (middle). Coomassie stained loading controls are shown (bottom).

Fig. 3. Sialic acid–dependent IgG anti-inflammatory activity is linkage specific, although attenuation of antibody-dependent cytotoxicity is not. (**A**) In vitro sialylated SNA⁺ IVIG Fc fragments were administered to mice 1 hour before K/BxN sera, and paw swelling was monitored over the next several days. Means and SDs (error bars) of four mice per group are plotted. *P < 0.05, as determined by ANOVA, followed by Tukey's post hoc test. (**B**) To determine whether the reduced antibody-dependent cell-mediated cytotoxicity of sialylated antibodies was also dependent on specific linkages, platelet-depleting 6A6-IgG2b antibodies were sialylated and administered to mice, and platelet counts were determined at 0, 4, and 24 hours after treatment. Means and SDs (error bars) of five mice per group are plotted; *P < 0.05, as determined by ANOVA, followed by Tukey's post hoc test.
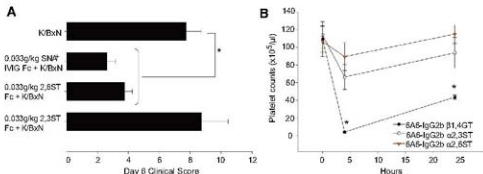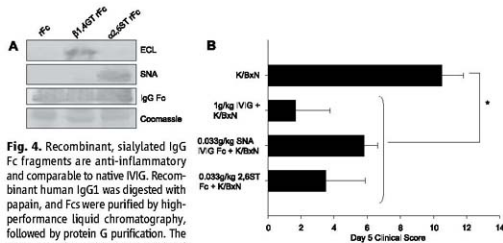
Fig. 4. Recombinant, sialylated IgG Fc fragments are anti-inflammatory and comparable to native IVIG. Recombinant human IgG1 was digested with papain, and Fcs were purified by high-performance liquid chromatography, followed by protein G purification. The rFcs were galactosylated and sialylated in vitro with $\alpha 2,6ST$. (**A**) Glycosylation was confirmed by lectin blotting for terminal galactose with ECL (top) or $\alpha 2,6$-sialic acid with SNA (middle); coomassie loading controls are shown (bottom). (**B**) Mice were administered 1 g/kg IVIG, 0.033 g/kg SNA$^+$ IVIG Fcs, or 0.33 g/kg sialylated rFc (2,6ST rFc) 1 hour before K/BxN sera, and footpad swelling was monitored over the next several days. Means and SDs (error bars) of clinical scores of four to five mice per group are plotted. *$P < 0.05$, as determined by Kruskal-Wallis ANOVA, followed by Dunn's post hoc test.

cytotoxic IgGs to engage activation FcRs and trigger inflammatory responses (*1, 2, 5, 24*). Along these lines, we have observed the preferential binding of 2,6-Fcs to marginal-zone macrophages, implicating a lectin expressed on these cells in this anti-inflammatory activity of IVIG. Furthermore, if our ability to recapitulate this anti-inflammatory property of IVIG in a recombinant molecule at a much reduced dose can be reproduced in the treatment of human autoimmune diseases, the studies described here will facilitate the development of this recombinant alternative as a broadly acting anti-inflammatory drug for the treatment of autoimmune disorders.

References and Notes
1. A. Samuelsson, T. L. Towers, J. V. Ravetch, *Science* **291**, 484 (2001).
2. P. Bruhns, A. Samuelsson, J. W. Pollard, J. V. Ravetch, *Immunity* **18**, 573 (2003).
3. Y. Kaneko, F. Nimmerjahn, M. P. Madaio, J. V. Ravetch, *J. Exp. Med.* **203**, 789 (2006).
4. Y. Kaneko, F. Nimmerjahn, J. V. Ravetch, *Science* **313**, 670 (2006).
5. F. Nimmerjahn, J. V. Ravetch, *J. Exp. Med.* **204**, 11 (2007).
6. N. Takahashi *et al.*, *Biochemistry* **26**, 1137 (1987).
7. Materials and methods are available as supporting material on Science Online.
8. A. S. Korganow *et al.*, *Immunity* **10**, 451 (1999).
9. M. Dalziel, I. McFarlane, J. S. Axford, *Glycoconj. J.* **16**, 801 (1999).
10. J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd, R. A. Dwek, *Annu. Rev. Immunol.* **25**, 21 (2007).
11. T. Hennet, D. Chui, J. C. Paulson, J. D. Marth, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4504 (1998).
12. J. N. Arnold *et al.*, *J. Immunol.* **173**, 6831 (2004).
13. H. Kitagawa, J. C. Paulson, *J. Biol. Chem.* **269**, 17872 (1994).
14. R. Jefferis *et al.*, *Biochem. J.* **268**, 529 (1990).
15. M. R. Wormald *et al.*, *Biochemistry* **36**, 1370 (1997).
16. B. J. Scallon, S. H. Tam, S. G. McCarthy, A. N. Cai, T. S. Raju, *Mol. Immunol.* **44**, 1524 (2007).
17. A. Varki, *Nature* **446**, 1023 (2007).
18. F. Lehmann, E. Tiralongo, J. Tiralongo, *Cell. Mol. Life Sci.* **63**, 1331 (2006).
19. T. M. Tumpey *et al.*, *Science* **315**, 655 (2007).
20. A. Ibricevic *et al.*, *J. Virol.* **80**, 7469 (2006).
21. F. G. Hanisch, J. Hacker, H. Schroten, *Infect. Immun.* **61**, 2108 (1993).
22. D. C. Mayer, O. Kaneko, D. E. Hudson-Taylor, M. E. Reid, L. H. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5222 (2001).
23. P. A. Orlandi, F. W. Klotz, J. D. Haynes, *J. Cell Biol.* **116**, 901 (1992).
24. F. Nimmerjahn, J. V. Ravetch, *Science* **310**, 1510 (2005).
25. We thank P. Smith, A. Miguelez, A. Magracheva, and J. Pagan for excellent technical support and J. Stavenhagen for providing the 4-20-20 construct. R.M.A. is an Irvington Institute fellow of the Cancer Research Institute. This work was performed with support from Centarus Pharmaceuticals and NIH to J.V.R. and by NIH grant GM60938 to J.C.P. J.V.R. is a founder and shareholder in Centarus Pharmaceuticals.

Reconstitution of Pilus Assembly Reveals a Bacterial Outer Membrane Catalyst

Mireille Nishiyama, Takashi Ishikawa, Helene Rechsteiner, Rudi Glockshuber*

Type 1 pili from uropathogenic *Escherichia coli* are a prototype of adhesive surface organelles assembled and secreted by the conserved chaperone/usher pathway. We reconstituted type 1 pilus biogenesis from purified pilus proteins. The usher FimD acted as a catalyst to accelerate the ordered assembly of protein subunits independently of cellular energy. Its activity was highly dependent on the adhesin subunit FimH, which triggered the conversion of FimD into a high-efficiency assembly catalyst. Furthermore, a simple kinetic model adequately rationalized usher-catalyzed pilus assembly in vivo. Our results contribute to a mechanistic understanding of protein-catalyzed biogenesis of supramolecular protein complexes at the bacterial outer cell membrane.

A dhesive pili are filamentous proteinaceous surface organelles that mediate adhesion and invasion of pathogenic bacteria to host tissues (*1, 2*). Type 1 pili, an important virulence factor in uropathogenic *Escherichia coli* strains (*3*), consist of a rigid helical rod composed of multiple copies of the major subunit,

FimA, and a short linear tip fibrillum formed by subunits FimF and FimG and the adhesin FimH (*4*) (Fig. 1A). FimH differs from all other subunits in that its pilin domain is linked to an amino-terminal mannose-binding lectin domain that recognizes the target receptor uroplakin Ia (*5, 6*). Biogenesis of type 1 pili is governed by the

conserved chaperone/usher pathway (*7, 8*). In the periplasm, the pilus chaperone, FimC, forms stoichiometric complexes with individual type 1 pilus subunits and directs them to the outer membrane usher, FimD. In FimC-subunit complexes, the incomplete immunoglobulin-like fold of the subunits is completed by a β-strand segment of FimC (*5, 9*). During pilus assembly, this segment of the chaperone is replaced by an amino-terminal extension of an incoming subunit in a reaction termed donor strand exchange (DSE) (*10–13*). The usher, FimD, orchestrates the entire assembly and secretion process at the outer membrane and is pivotal for pilus biogenesis. It uses an amino-terminal periplasmic domain, termed FimD$_N$, for the recognition and discrimination of incoming chaperone-subunit complexes (*14*). Left unexplored are the processes that follow this initial binding event. The potential function of the usher as catalyst in subunit assembly has been proposed (*7, 12, 13*) but never demonstrated experimentally.

Institute of Molecular Biology and Biophysics, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: rudi@mol.biol.ethz.ch

We have overproduced and purified FimD from *E. coli* outer membranes (fig. S1, A to D) and analyzed its ability to polymerize the major structural subunit, FimA, from purified FimA monomers complexed with FimC (*15*). Spontaneous assembly of FimA at 37°C proceeded very slowly, with a half-life of about 10 hours (Fig. 1B). Electron microscopic analysis revealed short pilus fragments with a relatively uniform length distribution and an average length of 35 nm (Fig. 1C). These assemblies were considerably shorter than a typical pilus rod in vivo (*4*). However, when FimD, pre-incubated with FimC-FimH (see below), was added in catalytic amounts to FimC-FimA, the FimC-FimA pool was completely depleted within 20 min (Fig. 1B). Moreover, nativelike pili with an average length of about 0.5 μm were generated (Fig. 1D and fig. S2). Efficient pilus assembly therefore requires no cellular components other than pilus subunits, FimC, and FimD, and both pilus assembly and secretion proceed independently of cellular energy. This agrees with the finding that neither nucleotide hydrolysis nor an electrochemical gradient is required for pilus formation in vivo (*16*).

The high acceleration of FimA polymerization was observed only when FimD was pre-incubated with an equimolar concentration of FimC-FimH. FimD alone accelerated FimA polymerization only marginally (Fig. 1B). Similar low activities were observed when FimD was pre-incubated with FimC-FimF, FimC-FimG, FimC-FimH$_P$ (FimH$_P$ denotes the isolated FimH pilin domain), the isolated FimH lectin domain, or a 1:1 mixture of FimC-FimH$_P$ and the lectin domain (Fig. 1B and figs. S3A and S4). Therefore, the ability to convert FimD into a high-efficiency assembly catalyst is restricted to FimC-FimH, and the FimH lectin domain is essential for the activation of FimD. These data rationalize previous observations that a FimH variant lacking the lectin domain fails to restore FimH deficiency in vivo (*17*).

In the absence of FimC-FimH, FimD was still capable of assembling a small number of long pili, some of which were markedly longer (>1.5 μm) than those assembled in the presence of FimC-FimH (Fig. 1, D and E). The main reaction product, however, was a large number of short pilus fragments, presumably generated by spontaneous FimA polymerization (Fig. 1E). A similar distribution of pilus lengths was found when FimA polymerization was followed in the presence of FimD pre-incubated with FimC-FimF, FimC-FimG, or FimC-FimH$_P$ (fig. S3, B and C). Thus, in the absence of full-length FimH, the conversion of FimD into a high-efficiency assembly catalyst proceeds only slowly, so that spontaneous polymerization of FimA stays dominant. Accordingly, *E. coli fimH* deletion strains make very few and abnormally long pili (*18*).

The kinetics of FimD activation by FimC-FimH followed clear first-order kinetics (Fig. 1F), demonstrating that the rate-limiting step in FimD activation is not the binding of FimC-FimH to FimD but rather a unimolecular process in the formed ternary complex. Presumably, this is associated with a conformational change in FimD that could be coupled to the insertion of the FimH lectin domain into the FimD translocation pore. A conformational transition in FimD is indeed observed in vivo upon co-expression of FimC-FimH with FimD as reflected in a decreased susceptibility of FimD to proteolysis (*19*).

We next pre-incubated FimD with different amounts of FimC-FimH before initiation of pilus rod assembly by the addition of excess FimC-FimA. The activity of FimD reached a maximum at a FimC-FimH to FimD ratio of 1:1 and then gradually decreased with increasing amounts of FimC-FimH (Fig. 1G). Excess FimC-FimH therefore competes with FimC-FimA in FimD-catalyzed pilus rod assembly. In fact, the initial chaperone-subunit binding site of FimD, FimD$_N$, has a substantially higher affinity for FimC-FimH than for FimC-FimA (*20*). In accordance with
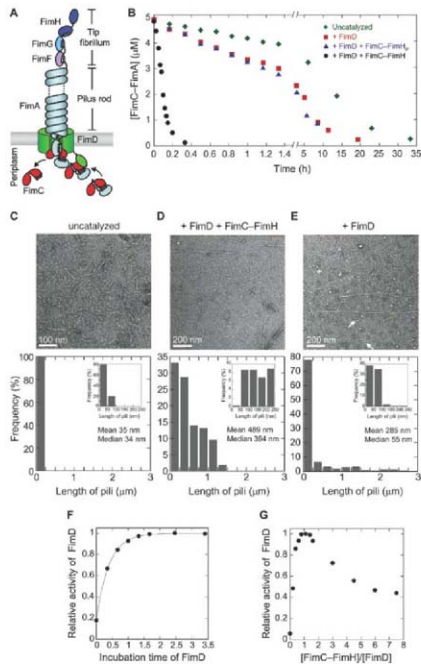
Fig. 1. The usher FimD catalyzes FimA polymerization. (**A**) Schematic of type 1 pilus assembly via the chaperone/usher pathway. (**B**) Kinetics of spontaneous and FimD-catalyzed polymerization of FimA at 37°C. Initial concentrations were 5 μM for FimC-FimA and 0.25 μM each for FimD, FimC-FimH, and FimC-FimH$_P$. (**C** to **E**) Electron micrographs (top) and length distribution histograms (bottom) of generated pili. Arrows in (E) indicate short pilus fragments coexisting with longer pili. (**F**) Activation of FimD by equimolar amounts of FimC-FimH. The solid line represents a fit to first-order kinetics. (**G**) Catalytic activity of FimD as a function of the molar FimC-FimH equivalents used for FimD activation.

this, overproduction of FimH in vivo results in nearly bald cells without pili [21].

An initial lag phase in FimD-catalyzed pilus rod assembly became apparent upon lowering

Fig. 2. Rate-limiting step in usher-catalyzed pilus rod assembly. **(A)** Kinetics of FimD-catalyzed FimA polymerization at 37° and 23°C after FimD activation by FimC-FimH (same setup as in Fig. 1B). **(B)** Length distribution histograms of generated pili after different reaction times at 23°C. **(C)** Influence of the subunits FimG and FimF on the kinetics of catalyzed FimA polymerization at 23°C. Continuous lines represent the global fit of the experimental data to the assembly model proposed in Fig. 3A. Schematics on the right depict the primed FimD molecules with the respective tip fibrillum subunits before initiation of pilus rod formation. **(D)** Length distribution histograms of pili formed at the end of the reactions in (C).

the temperature (Fig. 2A). Electron microscopic analysis of generated pili after different reaction times revealed that the distribution of pilus length significantly broadened over the course of the reaction, suggesting that different FimD molecules initiated pilus assembly at different times (Fig. 2B).

We hypothesized that the delay in the initiation of the assembly reaction was either due to a nucleation step similar to the assembly of other fibrous structures or arose from the slow incorporation of the first FimA subunit into the growing pilus, which, in the absence of the tip fibrillum subunits FimF and FimG, would involve a nonnatural interaction between FimH and FimA [4]. We thus tested whether the addition of FimC-FimF during the pre-incubation of FimD with FimC-FimH (at a FimD:FimF:FimH ratio of 1:1:1) diminished the lag phase. The lag phase was indeed significantly reduced, yet still existent, and FimC-FimA was depleted about four times faster compared with the reaction in which FimD was pre-incubated only with FimC-FimH (Fig. 2C and fig. S5A). FimC-FimG had a similar but less-pronounced effect than FimC-FimF (Fig. 2C and fig. S5A). These results provide strong evidence that the lag phase reflects the DSE reaction between the first incoming FimC-FimA complex and the proximal, chaperone-capped tip fibrillum subunit. Consequently, the rate-limiting step in catalyzed pilus rod assembly is the incorporation of the first FimA subunit into the preformed tip fibrillum, and this rate directly determines the number of FimD molecules that participate in rod formation. This is corroborated by length distribution analysis of generated pili, which revealed that more and shorter pili were formed in the presence of FimC-FimF, with a narrower distribution of pilus length (Fig. 2D and fig. S5, B to E). In fact, previous in vivo studies showed that the deletion of the *fimF* gene, in

Fig. 3. Mechanism of usher-catalyzed pilus rod assembly. **(A)** Model of FimD-catalyzed pilus rod formation after completion of the assembly of the tip fibrillum. Catalyzed rod formation comprises: (i) binding of FimC-FimA to FimD$_N$ with the rate constants k_{on} and k_{off} (steps 1 and 3), (ii) DSE between FimC-FimA and the last tip fibrillum subunit FimX (X stands for H, G, or F) capped by FimC, with the rate constant $k_{(A\rightarrow X)}$, and (iii) DSE between two FimC-FimA complexes, with the rate constant $k_{(A\rightarrow A)}$. Steps 3 and 4 are repeated until the FimC-FimA pool is depleted. **(B)** Ratios between the rate constants of catalyzed DSE between FimA and respective subunits and that between FimA and FimH. Rate constants obtained from the global fit to experimental data at 23°C are given in [24]. **(C)** Ratios between the rate constants of spontaneous DSE between FimA$_A$ and respective subunits and that between FimA$_A$ and FimH. Measured rate constants at 37°C are given in [25].
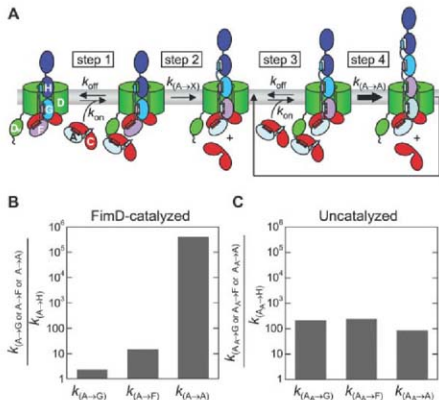
contrast to the deletion of the *fimG* gene, significantly reduces the number of pili per cell (22). Concurrent deletion of both genes further decreases the number of pili per cell, whereas pilus length increases (22). Furthermore, the kinetics of FimA assembly and the length distribution of generated pili were comparable when FimC-FimF was included alone or together with FimC-FimG (Fig. 2C and fig. S5, D and E). This result is expected because the proximal end of the tip fibrillum is formed by the same subunit in both cases, that is, FimF (compare with Fig. 2C). Given that all tip fibrillum subunits are present in the latter case, our data confirm that the tip fibrillum is assembled in the order FimH-FimG-FimF (4, 23).

On the basis of these results, we propose a simple model for usher-catalyzed pilus rod assembly (Fig. 3A). A global kinetic fit of our experimental data to this model revealed that the binding of the FimA subunit to FimF proceeds about 15-fold faster than the binding to FimH (all subunits in complex with the chaperone) (Fig. 3B) (24). With FimG as acceptor subunit for the first FimA subunit, the reaction was only twofold faster than the binding of FimA to FimH (24). Remarkably, the rate of DSE between two FimC-FimA complexes was about five orders of magnitude higher than that between FimC-FimA and any of the tip fibrillum subunits bound to the chaperone (Fig. 3B) (24). To test whether the preferred DSE reaction between two FimA molecules is mediated by the catalyst FimD, we determined the rate of spontaneous DSE between different acceptor subunits bound to FimC and the donor subunit, FimA$_A$ (Fig. 3C) (25). FimA$_A$ is a self-complemented FimA variant that is unable to self-polymerize but retains the ability to bind to another subunit via its amino-terminal donor strand (12). Analogous to the catalyzed reaction, the FimC-FimH complex exhibited the lowest DSE rate with FimA$_A$ (~100-fold lower compared with those of other FimC-subunit complexes) (25). However, a clear difference to the catalyzed reaction is that the other three subunits, each bound to the chaperone, underwent DSE with FimA$_A$ at similar rates (Fig. 3C) (25). Thus, the high DSE rate between two FimC-FimA complexes is the direct consequence of FimD catalysis. The catalytic power of the usher, therefore, does not primarily lie in accelerating the incorporation of the first FimA subunit into the preformed tip fibrillum, which occurs only once in the entire assembly process, but rather in the following polymerization reaction during which hundreds to thousands of FimA molecules are assembled.

Adhesive pili have gained increasing interest in recent years as attractive, hitherto unexploited targets for the development of novel antimicrobial drugs (26). Our in vitro reconstituted assembly system may not only serve as a tool for the identification of compounds that specifically block individual steps in pilus bio-genesis but also may provide a framework for a more detailed mechanistic understanding of chaperone/usher-dependent assembly and secretion of related pili from a wide variety of Gram-negative pathogens.

References and Notes
1. J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, G. Grandi, *Nat. Rev. Microbiol.* **4**, 509 (2006).
2. J. J. Martinez, M. A. Mulvey, J. D. Schilling, J. S. Pinkner, S. J. Hultgren, *EMBO J.* **19**, 2803 (2000).
3. I. Connell et al., *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9827 (1996).
4. E. Hahn et al., *J. Mol. Biol.* **323**, 845 (2002).
5. D. Choudhury et al., *Science* **285**, 1061 (1999).
6. G. Zhou et al., *J. Cell Sci.* **114**, 4095 (2001).
7. M. Vetsch et al., *Nature* **431**, 329 (2004).
8. S. D. Knight, *Adv. Exp. Med. Biol.* **74**, 2007 (2007).
9. F. G. Sauer et al., *Science* **285**, 1058 (1999).
10. F. G. Sauer, J. S. Pinkner, G. Waksman, S. J. Hultgren, *Cell* **111**, 543 (2002).
11. A. V. Zavialov et al., *Cell* **113**, 587 (2003).
12. M. Vetsch et al., *EMBO Rep.* **7**, 734 (2006).
13. H. Remaut et al., *Mol. Cell* **22**, 831 (2006).
14. M. Nishiyama et al., *EMBO J.* **24**, 2075 (2005).
15. Materials and methods are available as supporting material on *Science* Online.
16. F. Jacob-Dubuisson, R. Striker, S. J. Hultgren, *J. Biol. Chem.* **269**, 12447 (1994).
17. D. Munera, S. Hultgren, L. A. Fernandez, *Mol. Microbiol.* **64**, 333 (2007).
18. M. A. Schembri, L. Pallesen, H. Connell, D. L. Hasty, P. Klemm, *FEMS Microbiol. Lett.* **137**, 257 (1996).
19. E. T. Saulino, D. G. Thanassi, J. S. Pinkner, S. J. Hultgren, *EMBO J.* **17**, 2177 (1998).
20. M. Nishiyama, M. Vetsch, C. Puorger, I. Jelesarov, R. Glockshuber, *J. Mol. Biol.* **330**, 513 (2003).
21. P. Klemm, G. Christiansen, *Mol. Gen. Genet.* **208**, 439 (1987).
22. P. W. Russell, P. E. Orndorff, *J. Bacteriol.* **174**, 5923 (1992).
23. C. H. Jones et al., *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2081 (1995).
24. $k_{on} \geq 2 \times 10^5\ M^{-1}\ s^{-1}$, $k_{off} = 2\ s^{-1}$, $k_{(A \to F)} = 4 \times 10^{-5}\ s^{-1}$, $k_{(A \to G)} = 8 \times 10^{-6}\ s^{-1}$, $k_{(A \to H)} = 5 \times 10^{-4}\ s^{-1} \cdot k_{(A \to A)} = 16\ s^{-1}$. The rate constants k_{on} and k_{off} describe the association with and dissociation from FimD$_A$ of FimC-FimA, respectively, and the rate constants $k_{(A \to F)}$, $k_{(A \to G)}$, $k_{(A \to H)}$, and $k_{(A \to A)}$ describe the FimD-mediated DSE reactions between FimA as donor subunit and FimH, FimG, FimF, and FimA as acceptor subunit, respectively.
25. $k_{(A \to H)} = 3.0 \times 10^{-7} \pm 0.3 \times 10^{-7}\ M^{-1}\ s^{-1}$, $k_{(A \to G)} = 6.5 \pm 0.2\ M^{-1}\ s^{-1}$, $k_{(A \to F)} = 7.3 \pm 0.2\ M^{-1}\ s^{-1}$, $k_{(A \to A)} = 2.6 \pm 0.1\ M^{-1}\ s^{-1}$. The rate constants $k_{(A \to H)}$ describe the spontaneous DSE reaction between FimA$_A$ as donor subunit and the subunit FimH bound to FimC (that is FimH, FimG, FimF, or FimA).
26. J. S. Pinkner et al., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17897 (2006).
27. We thank K. Hollenstein and L. Thöny-Meyer for constructive discussions and comments on the manuscript and E. Weber-Ban for helpful discussions on kinetic data. We also acknowledge support of the Electron Microscopy Center (EMEZ) of the ETH Zurich. This work was supported by the Swiss National Science Foundation 3100A0-100787 (R.G.) and the ETH Zurich in the framework of the National Center for Competence in Research (NCCR) Structural Biology Program.

Structural Basis of Toll-Like Receptor 3 Signaling with Double-Stranded RNA

Lin Liu,[1] Istvan Botos,[1] Yan Wang,[2] Joshua N. Leonard,[2] Joseph Shiloach,[3] David M. Segal,[2] David R. Davies[1]*

Toll-like receptor 3 (TLR3) recognizes double-stranded RNA (dsRNA), a molecular signature of most viruses, and triggers inflammatory responses that prevent viral spread. TLR3 ectodomains (ECDs) dimerize on oligonucleotides of at least 40 to 50 base pairs in length, the minimal length required for signal transduction. To establish the molecular basis for ligand binding and signaling, we determined the crystal structure of a complex between two mouse TLR3-ECDs and dsRNA at 3.4 angstrom resolution. Each TLR3-ECD binds dsRNA at two sites located at opposite ends of the TLR3 horseshoe, and an intermolecular contact between the two TLR3-ECD C-terminal domains coordinates and stabilizes the dimer. This juxtaposition could mediate downstream signaling by dimerizing the cytoplasmic Toll interleukin-1 receptor (TIR) domains. The overall shape of the TLR3-ECD does not change upon binding to dsRNA.

The Toll-like receptor (TLR) family comprises 10 to 12 type I integral membrane receptor paralogs that recognize pathogen-associated molecular signatures and initiate inflammatory responses (1–3). TLR3 ectodomain (ECD) binds double-stranded RNA (dsRNA) (4), a viral replication intermediate, and recruits the adaptor protein TRIF to its cytoplasmic Toll interleukin-1 receptor (TIR) domain, thereby initiating a signaling cascade that results in the

secretion of type I interferons and other inflammatory cytokines (5). Although the role of TLR3 in controlling infections is not fully understood, it is clear that TLR3 nonredundantly contributes to the prevention of herpes simplex encephalitis in children (6). The structure of human (h) TLR3-ECD has been determined by two laboratories (7, 8) and consists of a solenoid with 23 leucine-rich repeats bent into a horseshoe shape, capped at each end by specialized structures known as

leucine-rich repeat N-terminal (LRR-NT) and C-terminal (LRR-CT) domains. TLR3-ECD binds dsRNA only at acidic pH (pH 6.5 and below) (9, 10), reflecting its endosomal location in most cell types (11). Although TLR3-ECD is monomeric in solution, it binds as dimers to dsRNA and requires a dsRNA length of at least 40 to 50 base pairs (bp) to bind a single dimer and induce signaling (10). However, the molecular basis for dimerization and signaling remains unknown. We have therefore isolated and crystallized a complex containing two mouse (m) TLR3-ECD molecules bound to a 46-bp dsRNA.

We first determined the monomeric ligand-free (apo)–mTLR3-ECD structure by molecular replacement with the hTLR3-ECD structure (table S1). mTLR3-ECD has 78% sequence identity with hTLR3-ECD (fig. S1), and, as expected, the structures of the two ECDs are also highly homologous. mTLR3-ECD superimposes with hTLR3-ECD with a root mean square deviation of 1.36 Å (for 659 Cα atoms, pdb code 2A0Z)

[1]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [2]Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. [3]Biotechnology Unit, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

*To whom correspondence should be addressed: david.davies@nih.gov

and 0.93 Å (for 613 Cα atoms, pdb code 1ZIW) (fig. S2). As in hTLR3-ECD, a horseshoe-shaped solenoid is formed by 23 LRRs, with N-terminal and C-terminal capping motifs. A large, parallel β sheet defines the concave surface, and two insertions in LRRs 12 and 20 protrude from the convex surface. Similar to hTLR3-ECD, the mTLR3-ECD molecule has a glycan-free surface. Both molecules contain 15 predicted N-linked glycosylation sites, but two of these are in different positions in the mouse and human proteins (fig. S1). Eleven glycan moieties are visible in the mTLR3-ECD electron density maps.

We next solved the structure of the TLR3 signaling complex, consisting of two mTLR3-ECD molecules bound to one 46-bp dsRNA oligonucleotide (Fig. 1) by using the apo-mTLR3-ECD structure for molecular replacement. This structure identifies multiple intermolecular contacts that stabilize the complex. The dsRNA interacts with both an N-terminal and a C-terminal site on the glycan-free surface of each mTLR3-ECD, which are on opposite sides of the dsRNA (Fig. 1A), with the C termini in contact (Fig. 1B) and the N termini outstretched at opposing ends of the linear dsRNA molecule. The length of the complex is 141 Å. The overall structure of mTLR3-ECD does not change upon binding to dsRNA, supporting a signaling mechanism in which ligand-induced receptor dimerization brings the two cytoplasmic TIR domains into contact, thus triggering a downstream signaling cascade. The dsRNA in the complex retains a typical A-DNA–like structure, in which the ribose-phosphate backbone and the position of the grooves are the major determinants of binding. The mTLR3-ECD interacts with the

sugar-phosphate backbones, but not with individual bases. This explains why TLR3 lacks specificity for any particular nucleotide sequence (4, 10).

The first dsRNA:TLR3 interaction site is located close to the C terminus, on LRR19 to LRR21 (table S2 and Fig. 2A). In the complex, these binding sites from two mTLR3-ECD monomers face each other across the dsRNA. Residues within contact distance of the RNA include Asn[515], Asn[517], His[539], Asn[541], and Arg[544], which are all well-conserved in vertebrates (fig. S3). Mutational analysis previously showed that His[539]Glu and Asn[541]Ala were inactive and failed to bind dsRNA, whereas Asn[517]Ala and Arg[544]Ala retained activity, indicating that these latter residues are not essential for binding (12).

The second dsRNA:TLR3 interaction site is located on the N-terminal end (LRR-NT to LRR3) of the glycan-free surface and is formed by residues His[39], His[60], Arg[64], Phe[84], Ser[86], His[108], and Glu[110] (table S2 and Fig. 2C). A notable feature of this site is the presence of three conserved residues: His[39], His[60], and His[108] (fig. S3), which are located in LRR-NT, LRR1, and LRR3, respectively. These residues appear to interact with consecutive phosphate groups on one dsRNA chain. In addition, less well-conserved residues Arg[64], Phe[84], Ser[86], and Glu[110] also interact with the ligand (table S2 and Fig. 2A). To test the functional importance of the three His residues, we mutated them to Ala or Glu and examined the ability of the mutant proteins to activate NF-κB. As seen in Fig. 2B, His[39]Ala and His[60]Ala are inactive, indicating that these residues are essential for dsRNA binding. In contrast, His[108]Ala, like His[539]Ala (12), retains activity, but mutation

Fig. 1. dsRNA:TLR3 signaling complex. Mouse TLR3 ectodomains (green and cyan) form a dimer on the dsRNA (blue and red). The N glycans are shown (light green and light blue). (**A**) The N- and C-terminal binding sites. (**B**) Illustration of how the two C-terminal domains are brought together in the complex. Figures generated with PyMol (DeLano Scientific, San Carlos, CA).
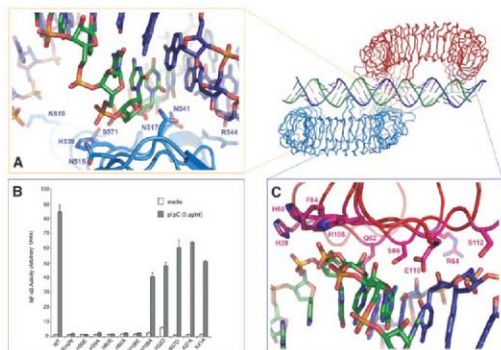
Fig. 2. The dsRNA binding sites of TLR3. (**A**) Residues involved in the interaction on the C-terminal site. (**B**) Illustrates the NF-κB activity of human TLR3 mutants stimulated with pI:pC. (**C**) Residues involved in binding at the N-terminal site. Each interaction involves two strands of the RNA. The dsRNA molecule may undergo a screw rotation of ±1 bp in the crystal.

to glutamate results in loss of function. These findings imply that His^{108} and His^{539} are not essential for ligand binding, but due to their proximity to the negatively charged phosphate groups in the dsRNA, mutation to a negatively charged glutamate disrupts ligand binding by electrostatic repulsion. Because the imidazole groups of His^{39} and His^{60} are required for TLR3 function, it is likely that the protonation of these residues at mildly acidic pH provides essential positive charges for ligand binding and accounts for the pH dependence of dsRNA binding to TLR3 (9, 10). The N-terminal dsRNA binding site is in the vicinity of a positively charged surface patch that was previously suggested as a potential dsRNA binding site (8). However, the residues from that surface patch are located on the convex face of the N terminus and do not come into direct contact with the dsRNA.

In the mTLR3-dsRNA 2:1 complex, the two LRR-CT domains are brought into proximity, forming a series of protein-protein interactions (table S3 and Fig. 3). Because these are the only interactions between the two mTLR3-ECDs, they must be responsible for coordinating the dimer on the dsRNA, thus facilitating the dimerization of the cytoplasmic TIR domains as suggested by Fig. 4. The distance between the centers of the LRR-CT domains is 26 Å. No other interactions exist to cause the two ECDs to dimerize on the dsRNA, and it will be interesting to learn whether similar interactions occur between LRR-CT domains in other TLRs. This protein-protein interaction also explains why TLR3-ECD was previously observed to bind to dsRNA with a high degree of cooperativity (10); although both the protein-protein interaction and

the binding of one TLR3-ECD to dsRNA are relatively weak, the dimeric complex is stabilized by multivalent intermolecular interactions.

Whereas most dsRNA contacts occur through residues on the glycan-free surface of mTLR3-ECD, the N-glycosyl moiety of Asn^{413}, which is located on the concave surface of mTLR3-ECD at LRR15, extends toward the dsRNA and directly contacts it. This contact may be more extensive in vivo, because mammalian glycan is typically larger than the insect glycan observed in our protein expression system. Although the Asn^{413} glycosylation site is well conserved (fig. S3), mutational analysis indicates that deletion of this glycan does not completely abrogate activity (12, 13).

Our results also help to explain some observations from a previous study by Takada et al. (14) in which single LRRs were sequentially deleted from hTLR3. Deletion of most LRRs located between the two dsRNA binding sites resulted in loss of function, which might be expected because such deletions would perturb the relative position of the two dsRNA binding sites on each TLR3-ECD. However, deletion LRRs 4, 11, and 17 did not abolish TLR3 function. The retention of function in these mutants may indicate that the TLR3 solenoid possesses sufficient flexibility such that it may compensate for the deletion and allow proper binding.

The dsRNA:TLR3 signaling complex features both similarities and differences with the structure of a complex containing TLR1 and TLR2 ECDs bridged by the ligand Pam_3CSK_4 (15). Two lipid chains of the ligand bind to a pocket in TLR2, and the third chain binds to a channel located at approximately the same region on TLR1. In both complexes, the ligand bridges two TLR molecules by the same glycan-free surface, resulting in the formation of a dimer in which the two TLR-ECDs are related by an ap-

proximate two-fold symmetry axis. TLR3 binds its ligand exclusively by surface contacts (mainly hydrogen bonding and electrostatic interactions), and although about the same surface area is buried in the TLR1:TLR2 interaction as in the TLR3 dimer, the intermolecular interactions involved differ substantially (table S4). The protein-protein interactions occur only at the LRR-CT in the TLR3:dsRNA complex, whereas the extensive interactions between TLR1 and TLR2 occur near the binding pockets. In the dsRNA:TLR3-ECD complex the two C-terminal residues are ~25 Å apart, whereas in the TLR1:TLR2 complex they are 42 Å apart. In the latter case, however, the native LRR-CT domains were replaced with a hagfish VLR LRR-CT. In both complexes the glycan-free surfaces are brought together by interaction with the ligands. For some TLRs, at least, a pattern may be emerging in which pathogen-associated ligands bind to TLRs by different mechanisms, but in each case, binding bridges two TLRs on the same glycan-free surface and forms dimers with similar overall architecture.

References and Notes

1. T. Kaisho, S. Akira, J. Allergy Clin. Immunol. 117, 979 (2006).
2. A. Iwasaki, R. Medzhitov, Nat. Immunol. 5, 987 (2004).
3. N. J. Gay, M. Gangloff, Annu. Rev. Biochem. 76, 141 (2007).
4. L. Alexopoulou, A. C. Holt, R. Medzhitov, R. A. Flavell, Nature 413, 732 (2001).
5. H. Oshiumi, M. Matsumoto, K. Funami, T. Akazawa, T. Seya, Nat. Immunol. 4, 161 (2003).
6. S. Y. Zhang et al., Science 317, 1522 (2007).
7. J. K. Bell et al., Proc. Natl. Acad. Sci. U.S.A. 102, 10976 (2005).
8. J. Choe, M. S. Kelker, I. A. Wilson, Science 309, 581 (2005).
9. O. de Bouteiller et al., J. Biol. Chem. 280, 38133 (2005).
10. J. N. Leonard et al., Proc. Natl. Acad. Sci. U.S.A. 105, 258 (2008).
11. M. Matsumoto et al., J. Immunol. 171, 3154 (2003).
12. J. K. Bell, J. Askins, P. R. Hall, D. R. Davies, D. M. Segal, Proc. Natl. Acad. Sci. U.S.A. 103, 8792 (2006).
13. J. Sun et al., J. Biol. Chem. 281, 11144 (2006).
14. E. Takada et al., Mol. Immunol. 44, 3633 (2007).
15. M. S. Jin et al., Cell 130, 1071 (2007).
16. We thank D. Xia for assistance with data collection and F. Dyda for valuable discussions. This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases and the National Cancer Institute and by an NIH/FDA Intramural Biodefense Award from the National Institute of Allergy and Infectious Diseases. We acknowledge the valuable assistance of the Protein Expression Laboratory, Science Applications International Corp., Frederick MD, in developing protein expression systems. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract W-31-109-Eng-38. The coordinates were deposited in the Protein DataBank, accession codes 3CIG (apo) and 3CIY (complex).
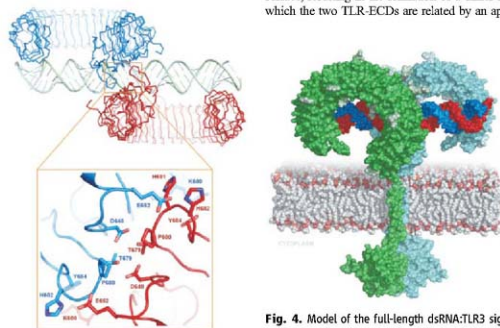
Fig. 3. Closeup of the C-terminal domain interacting residues. Some of these residues (678 to 681) are located on a conserved loop observed in other TLR structures.

Fig. 4. Model of the full-length dsRNA:TLR3 signaling complex. The proximity of the two C termini permits association of the transmembrane helices and the dimerization of the cytoplasmic TIR domains. The TIR domains were homology-modeled from the structure of the TLR10 TIR domains (pdb code 2J67).

Divergence of Quaternary Structures Among Bacterial Flagellar Filaments

Vitold E. Galkin,[1] Xiong Yu,[1] Jakub Bielnicki,[1] John Heuser,[2] Cheryl P. Ewing,[3] Patricia Guerry,[3] Edward H. Egelman[1]*

It has been widely assumed that the atomic structure of the flagellar filament from *Salmonella typhimurium* serves as a model for all bacterial flagellar filaments given the sequence conservation in the coiled-coil regions responsible for polymerization. On the basis of electron microscopic images, we show that the flagellar filaments from *Campylobacter jejuni* have seven protofilaments rather than the 11 in *S. typhimurium*. The vertebrate Toll-like receptor 5 (TLR5) recognizes a region of bacterial flagellin that is involved in the coiled-coil interactions in many other pathogenic bacteria, and this short region has diverged in *Campylobacter* and related bacteria, such as *Helicobacter pylori*, which are not recognized by TLR5. The driving force in the change of quaternary structure between *Salmonella* and *Campylobacter* may have been the evasion of TLR5.

Proteins with almost-unrecognizable sequence similarity can share highly conserved structures, suggesting that sequences can diverge more rapidly than structures over the course of evolution (*1*, *2*). Quaternary structure, on the other hand, may be much more sensitive than tertiary structure to small changes in sequence, and this sensitivity may provide a mechanism for large evolutionary divergence with only small changes in sequence. But relatively few examples exist for dramatic changes in quaternary structure (*3*, *4*). The bacterial flagellar filament has been a model system for understanding protein self-assembly, protein secretion, motility, and immunogenicity (*5*). The bacterial flagellar filament genes are highly conserved over large evolutionary distances, leading to the reasonable assumption that the atomic structure of the *Salmonella* flagellar filament (*6*) is a model for all homologous bacterial flagellar filaments (*7*). In fact, the assumption that a bacterium will have 11 protofilaments as in *Salmonella* has been taken as axiomatic (*8*). We have used electron microscopy (EM) to examine the flagellar filaments from *Campylobacter jejuni*, and we show that these contain only seven protofilaments.

C. jejuni is one of the major causes of bacterial diarrhea worldwide. The organism, which is a member of the epsilon division of the Proteobacteria, is characterized by a rapid, darting motility mediated by polar flagella. The filament is composed of a minor flagellin, FlaB, and a major flagellin, FlaA. However, full-length filaments can be formed when only FlaA is present, and these have nearly normal motility (*9*). *C. jejuni* flagellins, like those of other polar flagellates, are glycosylated, but, unlike other polar flagellates, glycosylation is required for filament formation in *Campylobacter* and the

related organism *Helicobacter pylori*. Flagellins from the epsilon Proteobacteria also are unique in that they fail to activate vertebrate Toll-like

receptor 5 (TLR5). In addition to mediating motility, which is essential for intestinal colonization, *Campylobacter* flagella, in the absence of type 3 secretion systems, also function to secrete nonflagellar proteins that modulate virulence, and the glycans on flagellin mediate auto-agglutination and microcolony formation (*10*).

The wild-type flagellar filaments from *Campylobacter* exist under normal conditions in different supercoiled states (Fig. 1A), consistent with the classical model of polymorphic switching between two discrete states of the component protofilaments (*11*, *12*). The two states are termed R (for right-handed) and L (for left-handed) on the basis of the twist of the filament when all protofilaments are in one state. In *Salmonella*, the structural parameters have been measured to very high precision, and the supercoiling arises from the fact that the R state of the protofilament is ~1.6% shorter than the L state (*13*). This non-equivalence of all protofilaments breaks the strict
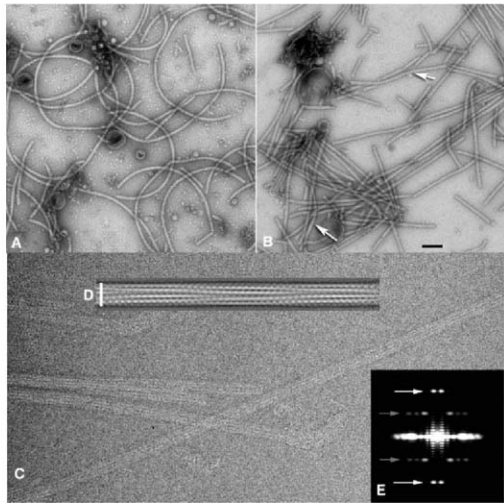
Fig. 1. Negatively stained and cryo-EMs of *Campylobacter* flagellar filaments. Negatively stained wild-type filaments (**A**) are curly, whereas the G508A mutant filaments (**B**) are rather straight. Nevertheless, some filaments can be seen [(B) arrows] that appear to be supercoiled. (**C**) A cryo-EM image of the G508A mutant shows long straight filaments. An averaged cryo-EM image of the filaments (**D**) was generated from 10,557 overlapping segments (each 200 pixels long, 2.4 Å per pixel) by using the IHRSR filrecon procedure. An averaged power spectrum (**E**), generated from 16,904 negatively stained segments each 460 Å long, shows two layer lines (marked with arrows). The higher layer line is at ~1/(26 Å), whereas the lower one is at ~1/(50 Å). A third layer line is so close to the equator that it cannot be resolved with segments of this length.

[1]Department of Biochemistry and Molecular Genetics, Box 800733, University of Virginia, Charlottesville, VA 22908–0733, USA. [2]Department of Cell Biology, Washington University School of Medicine, St. Louis, MO 63110, USA. [3]Enteric Diseases Department, Naval Medical Research Center, 503 Robert Grant Avenue, Silver Spring, MD 20910, USA.

*To whom correspondence should be addressed. E-mail: egelman@virginia.edu

helical symmetry of the flagellin subunits and leads to a curvature of the filament that generates thrust when the flagellar filament is rotated by the flagellar motor. The curvature of these flagellar filaments, as well as the heterogeneity of the protofilaments within them, make these wild-type filaments poor candidates for image analysis. The most detailed picture of the flagellar filament (6) has come from the Ala^{449}→Val^{449} (A449V) mutation (14, 15) in the coiled-coil domain D1 (fig. S1) that locks the protofilaments into the R state, forming straight filaments that cannot generate thrust. The corresponding mutation was constructed in the *flaA* gene of *C. jejuni* strain 81-176 and inserted into a chromosomal location in a flagellin-deletion background (16, 17). This mutation (A531V) does not eliminate motility as judged by swarming assays on agar plates, and electron micrographs of these filaments show that they are not straight and have similar supercoiling as wild-type filaments. This observation suggests that details of the packing of flagellin subunits must differ between *Salmonella* and *Campylobacter*.

We found that the Gly^{508} → Ala^{508} (G508A) mutation in *Campylobacter*, corresponding (fig. S1) to the G426A mutation in domain D1 that locks *Salmonella* flagellar filaments into the L state (14), produces largely straight filaments (Fig. 1, B and C) as judged by EM. Bacteria with the G508A mutation display no motility in solution by light-microscopic observation but do display partial motility in swarming assays using agar plates. The partial motility is not explained by revertants, because sequencing of

the flagellin gene from these apparently motile bacteria shows that the G508A mutation still exists and no other mutations (second site revertants) are present in the flagellin gene. Given the greater molecular weight of *Campylobacter* flagellin (59.5 kD and ~65 kD with glycosylation) compared with *Salmonella* flagellin (52.2 kD) and the fact that the sequence insertions in *Campylobacter* are in the D3 domain that is on the outside of the filament, it is surprising that the *Campylobacter* filaments are only ~180 Å in diameter (Fig. 1D), in contrast to the ~220 Å diameter of the *Salmonella* filaments. This is explained, as we show, by the different symmetry of these two filaments.

The straight filaments of *Salmonella* have a nearly crystalline order, which has allowed for structure determination by EM at near-atomic resolution (6). In contrast, images of the *Campylobacter* G508A filaments display limited diffraction in either negative stain or by cryo-EM, and we can show that this results in part from structural variability (movie S1). An averaged power spectrum (Fig. 1E) computed from these filaments shows only three layer lines.

We used quick freeze and deep etch (18) EM to visualize the surface features of the *Campylobacter* flagellar filaments (Fig. 2, A and C). Under the conditions used, all filaments examined had left-handed long-pitch helical stripes (Fig. 2D). However, the pitch of the long-handed stripes was variable. In contrast to the left-handed protofilaments in the quick freeze–deep etch images, cryo-EM images were

sorted into segments having either left-handed or right-handed protofilaments, and different twist states of either hand could be found (movie S1). Because the cryo-EM samples were prepared under different conditions (diluted in a low-salt buffer), we speculated that the low ionic strength and low salt used for cryo-EM imaging might introduce structural heterogeneity because changes in salt, pH, and ionic strength can change the polymorphic state of flagellar filaments (8, 19, 20). Under low ionic strength and low salt conditions, protofilaments could be seen by quick freeze–deep etch EM in a number of different states (Fig. 2, F to H), from L to nearly vertical to R. Interestingly, the transitions between these states appeared to be discrete and associated with "kinks" or bends (Fig. 2, F to H, arrows), which is probably similar to the highly cooperative transitions that have been observed in flagellar filaments at a macroscopic scale (21). Furthermore, many of the filaments were fragmented into short lengths because of the large forces that would result from a change in twist of a long filament adsorbed to the mica. For example, one end of a 1-μm-long segment would need to rotate by one or two full turns in the transition from the L state to the vertical state or the R state, respectively.

We used the iterative helical real space reconstruction (IHRSR) method (22) to generate reconstructions from the different twist states of the *Campylobacter* filaments imaged by cryo-EM. Resolution is limited by the heterogeneity of structural states, and we have not been able to do better than ~15 Å resolution. For com-

Fig. 2. Quick-freeze/deep etch EMs of the *Campylobacter* G508A flagellar filaments. (**A** and **C**) Filaments can be seen that contain protofilaments with a left-handed coil under normal buffer conditions. An averaged power spectrum (**B**) generated from 48 non-overlapping segments (each 1024 pixels long, with a sampling of 5.4 Å per pixel) shows a one-sided layer line at ~1/(50 Å) that arises from a left-handed three-start helix. A two-dimensional average (**D**) was made from 3806 overlapping segments (each 100 pixels long) that shows the long-pitch left-handed protofilaments. Although the metal shadowed images do not correspond to projections of a helical structure, the IHRSR technique (21) can be used to generate helical reconstructions from such images (4, 27), and a cross section from a reconstruction is shown (**E**). When these filaments are extensively washed with pure water after adsorption to the mica, structural heterogeneity can be seen (**F** to **H**). For all images, right-handed RecA-DNA filaments [labeled in (F)] were added as a control for the helical hand. In (F), the arrow indicates a transition from nearly vertical protofilaments (top of filament) to protofilaments with a left-handed coil (middle of filament). In (G), the arrow indicates a transition from nearly vertical protofilaments (left portion of filament) to a right-handed coil. In (H), the arrow indicates a transition from left-handed coiling (left portion of filament) to nearly vertical protofilaments. The scale bar in (F) indicates 500 Å.
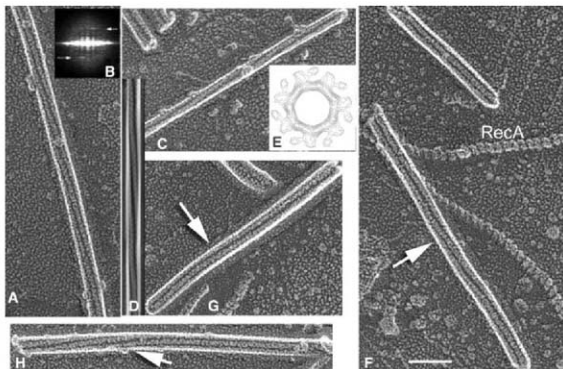
parison with the atomic model of *Salmonella* derived from right-handed filaments (Fig. 3, A and D), we show the R state of *Campylobacter* flagellar filaments (Fig. 3, B and E).

Despite the large difference in symmetry between *Salmonella* (~5.5 subunits per 26 Å pitch helical turn) and *Campylobacter* (~3.5 subunits per 26 Å pitch helical turn) flagellar filaments, there are some conserved parameters. In *Salmonella* the axial rise per subunit (13) within each of the 11 protofilaments is 52.7 Å in the L state and 51.9 Å in the R state, whereas in *Campylobacter* the axial rise per subunit within each of the 11 protofilaments is ~55.5 Å in the L state and ~52.5 Å in the R state. It should be noted that these values in *Salmonella* were determined to very high precision by using x-ray diffraction (13), whereas there is much greater error in our values using EM. In *Campylobacter* there is a ring of density surrounding the central lumen that contains seven rodlike features, in contrast to the 11 rodlike densities in *Salmonella*.

The length of these rodlike densities is very similar to the length of the coiled-coil D0 domain in *Salmonella* (Fig. 3C). The packing of these coiled-coils is looser in *Campylobacter* than in *Salmonella*, which is consistent with the nearly crystalline order in *Salmonella* and the structural heterogeneity in *Campylobacter*. Despite the fact that there are only seven coiled-coils surrounding the lumen in *Campylobacter* and 11 in *Salmonella*, the sizes of the lumens are comparable.

The resolution of our reconstruction is not sufficient to suggest where D1 is located. Residues 89 to 96 in *Salmonella* are located in D1, and these residues are recognized by TLR5 (23). Because mutation of these residues in *Salmonella* leads to a loss of bacterial motility (24), presumably by either interfering with flagellar assembly or preventing polymorphic switching, and these residues are not conserved in the comparison with *Campylobacter* (a bacterium that does not activate TLR5), it has been

suggested that the evolution of a flagellin that is both functional and evades TLR5 would require a complex sequence of mutations (23). Our results suggest that a substantially different packing of D1 domains between *Salmonella* and *Campylobacter* may have served this purpose.

Are *Campylobacter* flagellar filaments unique, or could other bacteria also have seven rather than 11 protofilaments? Previous observations in the literature may be explained by filament symmetries other than that found in *Salmonella*. It was shown (8) that the flagellar filaments of both *Idiomarina loihiensis* and *Pseudomonas aeruginosa* have different macroscopic forms than that found in *Salmonella*. In addition, fewer polymorphic states were observed for these filaments compared with those of *Salmonella*. It was shown by model building that a change of twist and curvature of the component protofilaments from the parameters found for *Salmonella* was needed to fit the observed macroscopic states of *I. loihiensis*, but the existence of 11 protofilaments in *Idiomarina* was taken as an axiom. An obvious prediction of the models for polymorphic switching (11, 12) is that the number of possible states is related to the number of protofilaments. With fewer protofilaments, we expect that *Campylobacter* would also have fewer polymorphic states than found in *Salmonella*.

One of the main mechanisms in evolutionary divergence is the accumulation of point mutations. Sequence divergence in proteins does not always map into corresponding structural divergence of these proteins because some protein folds have remained largely invariant while their sequences have diverged so greatly as to have no recognizable homology. For example, the core of the bacterial RecA recombination protein (25) can be superimposed quite well on the core of the mammalian F1 adenosine triphosphatase (ATPase) (26). However, RecA polymerizes to form a helical polymer, whereas this core is part of a hexameric ring in the F1 ATPase. We do not know at what point in the divergence of these two proteins that the change in quaternary structure took place. The results presented here show that, within a family of highly conserved proteins (flagellin), a different quaternary structure can be adopted with only limited sequence divergence of the D1 domain largely responsible for the helical symmetry. This may emerge as an important mechanism in the divergence of entire organisms.
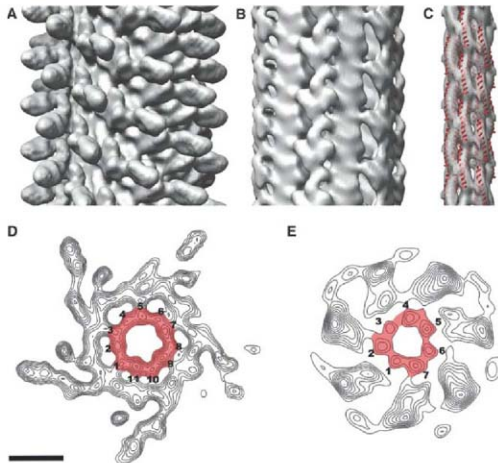
Fig. 3. Three-dimensional reconstructions of flagellar filaments. In (**A** and **D**) the atomic model of the *Salmonella* flagellar filament (6) has been filtered to 15 Å resolution for comparison with the cryo-EM reconstruction of the *Campylobacter* flagellar filament (**B** to **E**). The reconstruction has been generated from 10,557 overlapping segments (each 200 pixels long, 2.4 Å per pixel). (**C**) The core of the reconstruction (reconstruction truncated at a radius of 33 Å) shows vertical rods of density that are consistent with coiled-coils. The atomic model of the *Salmonella* D0 coiled-coil (red ribbons) has been fit to the *Campylobacter* reconstruction. Cross sections of the model (**D**) and the *Campylobacter* reconstruction (**E**) show the similar central lumen, surrounded by 11 protofilaments in *Salmonella* (D) and seven protofilaments in *Campylobacter* (E). The region surrounding the lumen, arising from D0 in *Salmonella*, has been colored red in (D) and (E). The scale bar for (D) and (E) indicates 50 Å.

References and Notes
1. L. A. Kelley, R. M. MacCallum, M. J. Sternberg, *J. Mol. Biol.* **299**, 499 (2000).
2. W. R. Pearson, M. L. Sierk, *Curr. Opin. Struct. Biol.* **15**, 254 (2005).
3. M. K. Frank, F. Dyda, A. Dobrodumov, A. M. Gronenborn, *Nat. Struct. Biol.* **9**, 877 (2002).
4. A. Orlova *et al.*, *Nat. Struct. Mol. Biol.* **14**, 921 (2007).
5. H. C. Ramos, M. Rumbo, J. C. Sirard, *Trends Microbiol.* **12**, 509 (2004).
6. K. Yonekura, S. Maki-Yonekura, K. Namba, *Nature* **424**, 643 (2003).
7. S. A. Beatson, T. Minamino, M. J. Pallen, *Trends Microbiol.* **14**, 151 (2006).

8. S. Shibata, M. Alam, S. Aizawa, *J. Mol. Biol.* **352**, 510 (2005).
9. P. Guerry, R. A. Alm, M. E. Power, S. M. Logan, T. J. Trust, *J. Bacteriol.* **173**, 4757 (1991).
10. P. Guerry, *Trends Microbial.* **15**, 456 (2007).
11. C. R. Calladine, *Nature* **255**, 121 (1975).
12. S. Asakura, *Adv. Biophys.* **1**, 99 (1970).
13. I. Yamashita *et al.*, *Nat. Struct. Biol.* **5**, 125 (1998).
14. H. C. Hyman, S. Trachtenberg, *J. Mol. Biol.* **220**, 79 (1991).
15. S. Kanto, H. Okino, S. Aizawa, S. Yamaguchi, *J. Mol. Biol.* **219**, 471 (1991).
16. S. Goon *et al.*, *Infect. Immun.* **74**, 769 (2006).
17. Materials and methods are available on *Science* Online.
18. J. Heuser, *Methods Cell Biol.* **22**, 97 (1981).
19. R. Kamiya, S. Asakura, *J. Mol. Biol.* **108**, 513 (1976).
20. R. Kamiya, S. Asakura, *J. Mol. Biol.* **106**, 167 (1976).
21. H. Hotani, *J. Mol. Biol.* **156**, 791 (1982).
22. E. H. Egelman, *Ultramicroscopy* **85**, 225 (2000).
23. E. Andersen-Nissen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9247 (2005).
24. K. D. Smith *et al.*, *Nat. Immunol.* **4**, 1247 (2003).
25. R. M. Story, I. T. Weber, T. A. Steitz, *Nature* **355**, 318 (1992).
26. J. P. Abrahams, A. G. Leslie, R. Lutter, J. E. Walker, *Nature* **370**, 621 (1994).
27. J. D. Woodward *et al.*, *J. Struct. Biol.* **161**, 111 (2008).
28. This work was supported by NIH grants EB001567 (E.H.E.) and AI043559 (P.G.). The electron microscopic reconstruction has been deposited in the Electron Microscopy Data Base with accession code EMD-5007.

The Antidepressant Fluoxetine Restores Plasticity in the Adult Visual Cortex

José Fernando Maya Vetencourt,[1]* Alessandro Sale,[1] Alessandro Viegi,[1] Laura Baroncelli,[1] Roberto De Pasquale,[1] Olivia F. O'Leary,[3] Eero Castrén,[3] Lamberto Maffei[1,2]

We investigated whether fluoxetine, a widely prescribed medication for treatment of depression, restores neuronal plasticity in the adult visual system of the rat. We found that chronic administration of fluoxetine reinstates ocular dominance plasticity in adulthood and promotes the recovery of visual functions in adult amblyopic animals, as tested electrophysiologically and behaviorally. These effects were accompanied by reduced intracortical inhibition and increased expression of brain-derived neurotrophic factor in the visual cortex. Cortical administration of diazepam prevented the effects induced by fluoxetine, indicating that the reduction of intracortical inhibition promotes visual cortical plasticity in the adult. Our results suggest a potential clinical application for fluoxetine in amblyopia as well as new mechanisms for the therapeutic effects of antidepressants and for the pathophysiology of mood disorders.

Clinically used antidepressant drugs (ADs) increase extracellular serotonin and/or norepinephrine levels, but the relationship between acute increases in these neurotransmitters and the clinical antidepressant effect, which develops with a time delay of several weeks, has remained unclear (*1*, *2*). Chronic antidepressant administration promotes neurogenesis and synaptogenesis in the adult hippocampus (*3*, *4*) as well as increased expression of the neurotrophin brain-derived neurotrophic factor (BDNF) and its primary receptor, TrkB (*5*, *6*). These cellular and molecular events seem to be necessary to mediate the therapeutic effects of ADs. The behavioral response to ADs is blocked if the induced neurogenesis is disrupted (*7*), whereas direct infusion of BDNF into the hippocampus or the overexpression of its receptor in transgenic mice induces an antidepressant response (*8*, *9*). Although neurogenesis, synaptogenesis, and BDNF signaling are events that correlate with neuronal plasticity, it remains to be demonstrated whether ADs induce functional modifications of neuronal circuitries in the brain.

We investigated whether chronic treatment with fluoxetine, a selective serotonin reuptake inhibitor (SSRI), restores plasticity in the adult visual system of the rat. We used two classical models of plasticity: (i) the ocular dominance (OD) shift of visual cortical neurons after monocular deprivation (MD) and (ii) the recovery of visual functions in the adult after long-term MD (amblyopia). These two plastic phenomena in the rat are restricted to a critical period during postnatal development and are absent in the adult because of a decline of plasticity that has been largely attributed to the maturation of intracortical inhibition (*10*, *11*).

We initially examined the effects of 1 week of MD on the OD plasticity of adult animals chronically treated with fluoxetine by recording visual evoked potentials (VEPs) in the binocular region of the primary visual cortex contralateral to the deprived eye. VEPs represent the integrated response of a population of neurons to patterned visual stimuli and are routinely used to evaluate visual acuity (VA) and binocularity alterations (*10*, *12*, *13*). OD plasticity assessed by VEPs in the rat follows a critical period, which ends by postnatal day 55 (P55) (*14*). We assessed OD (binocularity) calculating the contralateral-to-ipsilateral (C/I) VEP ratio; that is, the ratio of VEP amplitudes recorded by stimulating the eye contralateral and ipsilateral, respectively, to the

visual cortex where recording is performed. The C/I VEP ratio is around 2.5 in adult animals, reflecting the predominance of crossed fibers in rat retinal projections. MD in control adult animals (P100) did not change binocularity in the visual cortex contralateral to the deprived eye (C/I VEP ratio 2.73 ± 0.2, *n* = 5 animals) (Fig. 1A). In contrast, fluoxetine-treated adult rats showed a marked OD shift in favor of the nondeprived eye after MD (C/I VEP ratio 1.0 ± 0.08, *t* test *P* < 0.001, *n* = 5), thus displaying a plastic modification normally restricted to the early stages of brain development.

Changes of binocularity after MD may be due to an initial depression of the response to stimulation of the deprived eye, followed by an increased open-eye response (*15*), or only to an increase in the open eye's strength (*16*). To analyze which of these two components contributes to the reinstatement of OD plasticity induced by chronic antidepressant treatment in the adult rat, we compared VEP amplitudes in response to stimulation of either eye, respectively, in fluoxetine-treated and control animals. A reduction of the response to stimulation of the deprived eye after 1 week of MD (*t* test *P* < 0.05, *n* = 7) (fig. S1) was evident in fluoxetine-treated rats. No difference in the open eye's strength was observed after antidepressant treatment.

To exclude the possibility that pharmacological treatment per se affects binocularity in normal (not deprived) adult animals, we next assessed OD in rats with binocular vision after chronic treatment with fluoxetine. The C/I VEP ratio in adult animals after chronic antidepressant treatment did not differ from that observed in non–drug-treated rats (*t* test *P* = 0.703, *n* = 4) (fig. S2). To ensure that basic response properties of visual cortical neurons were not altered by chronic administration of fluoxetine, we analyzed orientation selectivity and cell responsiveness by recording single units in a subset (*n* = 2) of the same animals (*n* = 39 cells for fluoxetine; *n* = 55 cells for controls). Orientation selectivity of cortical neurons (the percentage of orientation-biased cells: 36%) was in the range of normal adult rats. No difference in cell responsiveness (*t* test *P* = 0.677) (fig. S3) was found between fluoxetine-treated and control animals.

We next evaluated the recovery of visual functions in adult amblyopic animals. Rats that were rendered amblyopic by long-term MD were

[1]Scuola Normale Superiore, Piazza dei Cavalieri, I-56100 Pisa, Italy. [2]Institute of Neuroscience, Consiglio Nazionale delle Recerche, Via Moruzzi 1, I-56100 Pisa, Italy. [3]Neuroscience Centre, University of Helsinki, 00014 Helsinki, Italy.

*To whom correspondence should be addressed. E-mail: jf.maya@in.cnr.it

chronically treated with fluoxetine, and the previously closed eye was opened while the open eye was closed (by reverse suture) during the last 2 weeks of antidepressant treatment (17). We measured VA by recording VEPs from the visual cortex contralateral to the long-term–deprived eye. In control animals, VA of the formerly deprived eye did not show any sign of recovery (0.62 ± 0.06 cycle/deg) as compared with the fellow eye (1.06 ± 0.01 cycle/deg) (Fig. 1B). In contrast, fluoxetine-treated adult rats showed a complete rescue of VA (0.97 ± 0.04 cycle/deg). A behavioral measure of VA (17), performed in the same animals before VEP recordings, confirmed the electrophysiological data: Complete recovery of VA (0.88 ± 0.02 cycle/deg) was evident in fluoxetine-treated long-term–deprived rats but not in controls (Fig. 1C). In the same animals in which VA was assessed, we also evaluated OD. In control animals, there was no rescue of binocularity in the visual cortex contralateral to the formerly deprived eye (C/I VEP ratio 1.11 ± 0.20, $n = 5$) (Fig. 1D), whereas fluoxetine-treated adult rats showed a full recovery of binocularity with a C/I ratio of 2.25 ± 0.17.

Because there is evidence that the maturation of intracortical inhibitory circuitries causes the end of plasticity in the visual system (18), we used in vivo brain microdialysis to investigate whether the visual cortical plasticity induced by fluoxetine was accompanied by decreased γ-aminobutyric acid–mediated (GABAergic) transmission. Extracellular basal levels of GABA were substantially reduced in the visual cortex of fluoxetine-treated adult rats (Fig. 2A) as compared with controls [two-way analysis of variance (ANOVA) repeated measures $P = 0.02$, post hoc Holm-Sidak test $P < 0.02$, $n = 5$]. No difference in extracellular glutamate levels was detected between fluoxetine-treated and control animals (two-way ANOVA repeated measures $P = 0.494$, $n = 5$) (fig. S4). The reduced levels of GABA in fluoxetine-treated animals did not affect spontaneous activity (t test $P = 1.0$) (fig. S5), which was assessed in the same animals in which basic response properties of cortical neurons were examined.

To assess the reduction of intracortical inhibition at the functional level, we next examined long-term potentiation (LTP) of layer II-III field potentials induced by θ-burst stimulation (TBS) of the white matter (WM-LTP) in the visual cortex, a form of synaptic plasticity that is absent in the adult because of the maturation of intracortical inhibitory circuitries (19). WM-LTP was fully restored in fluoxetine-treated adult rats (Fig. 2B). No WM-LTP was present in control animals. We next investigated long-term depression (LTD) of layer II-III field potentials after

low-frequency stimulation from the white matter, after chronic antidepressant treatment. We found that LTD was saturated in the visual cortex of MD adult rats chronically treated with fluoxetine as compared with controls [95.4% of pre-low-frequency stimulation baseline amplitude for MD fluoxetine-treated animals and 79.4% for controls].

Because chronic antidepressant administration increases the expression of the neurotrophin BDNF in limbic structures, most notably in the hippocampus (5, 20), we measured BDNF protein levels, using the enzyme-linked immunosorbent assay (ELISA) method, in the adult rat visual cortex after chronic fluoxetine administration. We found increased expression of BDNF in the visual cortex of fluoxetine-treated adult rats (t test $P < 0.04$, $n = 6$) (Fig. 2C). BDNF protein expression was similarly enhanced in the hippocampus (t test $P < 0.01$, $n = 6$) (fig. S6). To examine whether the increased BDNF expression is causally linked to the reinstatement of plasticity in the adult visual system, we then assessed OD in rats that were intracortically infused (via osmotic minipumps) with BDNF (1 ng µl^{-1}) in parallel to MD. Control animals infused with vehicle solution showed no change of binocularity in the visual cortex contralateral to the deprived eye (C/I VEP ratio 2.44 ± 0.1, $n = 2$). In contrast, adult rats intracortically infused with BDNF showed an OD shift in response to MD (C/I VEP ratio 1.32 ± 0.08 t test $P < 0.001$, $n = 4$) (Fig. 2D).

Finally, to directly test whether the reduction of intracortical inhibition underlies the reopening of visual cortical plasticity in adulthood, we evaluated OD in fluoxetine-treated adult rats that were infused intracortically with the benzodiazepine agonist diazepam (Dz) (2 mg ml^{-1}) or vehicle solution during the period of MD (Fig. 3A). Cortical Dz administration in adult rats chronically treated with fluoxetine totally prevented the OD shift induced by MD (Fig. 3B). Control animals intracortically infused with vehicle solution showed an OD shift in favor of the nondeprived eye after MD (C/I VEP ratio 1.07 ± 0.04, t test $P = 0.01$, $n = 3$).

Our findings demonstrate that chronic fluoxetine administration reinstates a juvenile-like form of OD plasticity in adulthood, which is indicated by a decrease in the response to stimulation of the deprived eye and promotes a complete recovery of visual functions in adult amblyopic rats. These effects are accompanied by a marked reduction of GABAergic inhibition and increased BDNF protein expression in the visual cortex. Furthermore, we provide evidence that the reduction of intracortical inhibition induced by ADs is a critical cellular mechanism to restore plasticity in the adult, because cortical infusion of Dz in fluoxetine-treated rats prevented the OD shift of cortical neurons after MD.

The effects induced by fluoxetine in adult visual cortical plasticity are surprisingly similar to those caused by environmental enrichment, a
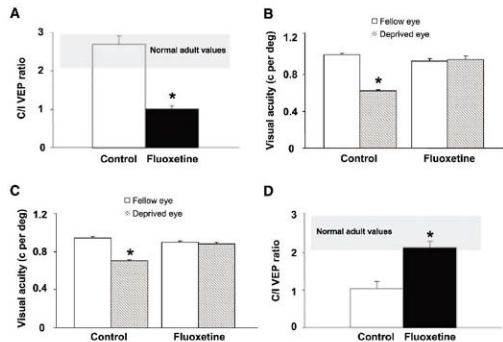
Fig. 1. Reactivation of visual cortical plasticity in adulthood after chronic treatment with fluoxetine. (**A**) MD in fluoxetine-treated adult rats induced an OD shift of visual cortical neurons in favor of the nondeprived eye (C/I VEP ratio 1.0 ± 0.08, t test $P < 0.001$, $n = 5$) but not in control animals (C/I VEP ratio 2.73 ± 0.2, $n = 5$). (**B**) and (**C**) VA recovery in adult amblyopic rats chronically treated with fluoxetine. VA of the formerly deprived eye assessed both electrophysiologically (B) and behaviorally (C) was lower than that of the fellow eye in control animals [paired t test $P < 0.001$ for (B) and $P < 0.001$ for (C), $n = 5$] but not in fluoxetine-treated adult rats [paired t test $P = 0.703$ for (B) and $P = 0.354$ for (O, $n = 5$]. (**D**) OD (binocularity) recovery in adult amblyopic rats after chronic treatment with fluoxetine. The C/I VEP ratio was significantly higher (t test $P < 0.002$, $n = 5$) in the visual cortex of fluoxetine-treated adult rats (C/I VEP ratio 2.25 ± 0.17) than in controls (C/I VEP ratio 1.11 ± 0.20, $n = 5$), and was in the range of adult animals with normal vision. Error bars represent SEM; asterisk indicates statistical significance.

condition characterized by increased exploratory behavior and sensory-motor stimulation, which we recently found to promote amblyopia recovery in adulthood through a reduction of intracortical inhibition (21). Our data suggest that the enhanced serotonergic transmission induced by chronic treatment with fluoxetine promotes functional and/or structural mechanisms that shift the intracortical inhibitory-excitatory balance, triggering plasticity in the adult visual cortex. We propose that the reduced GABAergic inhibition induced by chronic fluoxetine administration and increased BDNF expression open the pathway to the genes that regulate plasticity, thus allowing a functional modification of neuronal circuitries that underlies the sensitivity to MD in the adult and recovery from amblyopia. Because previous studies have shown that serotonin enhances neuronal responses to the excitatory aminoacid agonist N-methyl-D-aspartate in neocortical slices of adult rats (22), we cannot rule out the possibility that reinstatement of plasticity in the adult visual cortex induced by chronic fluoxetine administration may also involve an increased glutamatergic transmission.

We also observed the occurrence of WM-LTP in the visual cortex of fluoxetine-treated rats, a phenomenon that is usually absent in the adult but can be restored if GABA-mediated inhibition is reduced (23). Chronic administration of ADs in healthy humans increases the amplitude of the P1 and N1 components of VEPs in response to repeated presentation of visual stimuli, which has been suggested to be a form of synaptic plasticity (24). Taken together, these findings indicate that a similar increase of plasticity, which is described here in the adult visual cortex, may also take place in the human
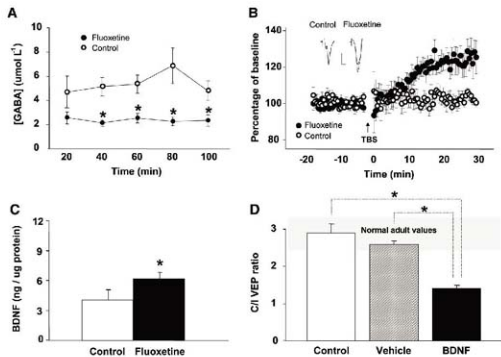
Fig. 2. Reduced intracortical inhibition and increased expression of BDNF in the adult rat visual cortex after chronic antidepressant treatment. (A) GABAergic neurotransmission in the visual cortex of fluoxetine-treated adult rats. In vivo brain microdialysis revealed that basal extracellular levels of GABA were significantly lower in fluoxetine-treated animals than in control rats (two-way ANOVA repeated measures $P = 0.02$, post hoc Holm-Sidak test $P < 0.02$ where indicated, $n = 5$). (B) LTP of neural transmission in the visual cortex. WM-LTP, measured 20 to 30 min after TBS, was significantly higher in the visual cortex of fluoxetine-treated animals than in controls (two-way ANOVA repeated measures $P < 0.005$, post hoc Student-Newman-Keuls test $P < 0.01$). Scale bars are 50% of baseline amplitude and 5 ms. (C) BDNF protein levels after antidepressant treatment. BDNF protein expression, quantified by means of ELISA, was significantly higher in the visual cortex of adult rats chronically treated with fluoxetine than in controls (t test $P < 0.04$, $n = 6$) of adult rats chronically treated with fluoxetine than in controls. (D) Intracortical administration of BDNF. Adult animals cortically infused with BDNF showed an OD shift in the visual cortex contralateral to the deprived eye as compared with control animals (t test $P < 0.001$, $n = 4$). Error bars represent SEM; asterisk indicates statistical significance.

Fig. 3. Cortical administration of Dz prevented the restoration of OD plasticity induced by chronic fluoxetine administration. (A) Schematic diagrams of the experimental procedure followed (top) and of the osmotic minipump implant and recording site of VEPs in the binocular visual cortex contralateral to the deprived eye (bottom). Cortical administration of the benzodiazepine agonist Dz was performed in parallel with MD during the last week of antidepressant treatment. Oc1B, primary visual cortex. (B) Blockade of OD plasticity in fluoxetine-treated rats intracortically infused with Dz. The C/I VEP ratio in the visual cortex contralateral to the deprived eye after MD in fluoxetine-treated adult animals that were cortically infused with the benzodiazepine agonist Dz (Fluox + Dz) was not different from that of control (nondeprived) animals (C/I VEP ratio 2.48 ± 0.29, t test $P = 0.483$, $n = 4$),

but differed significantly from either that of adult rats chronically treated with fluoxetine (Fluox) (t test $P = 0.001$, $n = 5$) or of animals cortically infused with vehicle solution (Fluox + Veh) (t test $P = 0.01$, $n = 3$). Error bars represent SEM; asterisk indicates statistical significance.

visual system during chronic SSRI treatment. Moreover, our results open the possibility that chronic treatment with antidepressants may also have similar effects in other brain areas, such as those involved in mood regulation in depressed patients, and suggest new mechanisms for the therapeutic effects induced by ADs and the pathophysiology of mood disorders.

Our finding that fluoxetine, a widely prescribed AD in humans, restores plasticity in the adult visual system suggests that antidepressants may be used as a complementary treatment to current therapies for human amblyopia. Although amblyopia can be prevented by occlusion treatment during childhood, there are currently no treatment strategies that would restore vision to the amblyopic eye in adults (25). Our data also indicate a potential clinical application for antidepressants in neurological disorders in which synaptic plasticity is compromised because of excessive intracortical inhibition (26–28).

References and Notes

1. E. J. Nestler, *Biol. Psychiatry* **44**, 526 (1998).
2. E. Castren, *Nat. Rev. Neurosci.* **6**, 241 (2005).
3. J. E. Malberg, A. J. Eisch, E. J. Nestler, R. S. Duman, *J. Neurosci.* **20**, 9104 (2000).
4. T. Hajszan, N. J. MacLusky, C. Leranth, *J. Neurosci.* **21**, 1299 (2005).
5. M. Nibuya, S. Morinobu, R. S. Duman, *J. Neurosci.* **15**, 7539 (1995).
6. E. Castren, *Curr. Opin. Pharmacol.* **4**, 58 (2004).
7. L. Santarelli *et al.*, *Science* **301**, 805 (2003).
8. Y. Shirayama, A. Chen, S. Nakagawa, D. Russell, R. S. Duman, *J. Neurosci.* **22**, 3251 (2002).
9. E. Koponen *et al.*, *Cell. Mol. Neurobiol.* **25**, 973 (2005).
10. Z. J. Huang *et al.*, *Cell* **98**, 739 (1999).
11. M. Fagiolini, T. K. Hensch, *Nature* **404**, 183 (2000).
12. V. Porciatti, T. Pizzorusso, L. Maffei, *Vision Res.* **39**, 3071 (1999).
13. H. Y. He, W. Hodos, E. M. Quinlan, *J. Neurosci.* **26**, 2951 (2006).
14. E. Guire, M. Lickey, B. Gordon, *J. Neurophysiol.* **81**, 121 (1999).
15. M. Y. Frenkel, M. F. Bear, *Neuron* **44**, 917 (2004).
16. N. B. Sawtell *et al.*, *Neuron* **38**, 977 (2003).
17. Materials and methods are available as supporting material on *Science* Online.
18. T. K. Hensch, *Nat. Rev. Neurosci.* **6**, 877 (2005).
19. A. Kirkwood, M. F. Bear, *J. Neurosci.* **14**, 1634 (1994).
20. C. D'Sa, R. S. Duman, *Bipolar Disord.* **4**, 183 (2002).
21. A. Sale *et al.*, *Nat. Neurosci.* **10**, 679 (2007).
22. J. N. Reynolds, A. Baskys, P. L. Carlen, *Brain Res.* **456**, 286 (1988).
23. A. Artola, W. Singer, *Nature* **330**, 649 (1987).
24. C. Normann, D. Schmitz, A. Furmaier, C. Doing, M. Bach, *Biol. Psychiatry* **62**, 373 (2007).
25. J. M. Holmes, M. P. Clarke, *Lancet* **367**, 1343 (2006).
26. V. S. Dani *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12560 (2005).
27. J. L. Rubenstein, M. M. Merzenich, *Genes Brain Behav.* **2**, 255 (2003).
28. F. Fernandez *et al.*, *Nat. Neurosci.* **10**, 411 (2007).
29. We thank L. Gianfranceschi for her valuable assistance with electrophysiological recordings. This work was supported by grants from Ministero dell'Università e della Ricerca, Programmi di Ricerca di Rilevante Interesse Nazionale, Fondo Integrativo Speciale Ricerca, Sigrid Jusélius Foundation, and the Academy of Finland and Center of Excellence in Molecular Neurosciences.

New Products Focus: Microscopy

Vibration-Isolation Workstations

Sturdy, versatile, and ergonomic, this line of vibration-isolation workstations features automatic leveling and low natural frequencies for excellent vertical and horizontal isolation efficiencies. The workstations can be configured for a wide variety of applications in which external vibrations could adversely affect the operation of sensitive equipment. The Standard High-Performance Workstation is the choice for most applications. A Heavy/Moving-Load Workstation is available to isolate loads of up to 2,800 pounds, including equipment with high centers of gravity. The Personal Workstation is an economical solution for small labs and other tight spaces. A Variable-Height Workstation allows the user to raise or lower the tabletop to accommodate the height of the user. The Cleanroom Workstation is designed for use in Class 1 and Class 10 cleanrooms. The MK26 Workstation makes use of the Minus K stiff spring and negative-stiffness mechanism to achieve low net vertical stiffness, ultralow natural frequencies, and high internal structural frequencies without affecting the static load supporting capability.

Kinetic Systems
For information 617-522-8700
www.kineticsystems.com

Supercontinuum Confocal Microscope

The Leica TCS SP5 X confocal microscope with supercontinuum laser provides freedom and flexibility that were unattainable until now in fluorescence microscopy. The supercontinuum laser allows the researcher to choose any excitation line within the continuous range of 470 nm to 670 nm. The Leica TCS SP5 X can precisely adapt to any fluorescent dye. Adjustment of the excitation line to the sample in 1 nm increments reduces cross-excitation and minimizes sample damage. The system allows multicolor excitation with up to eight excitation lines simultaneously. The patented Spectral Detection Technology provides the flexibility to choose any detection area in up to five spectral confocal photomultiplier tube channels. The system includes the acousto-optical beamsplitter, a Leica invention that provides high sensitivity and better transmission than any other dichroic mirror.

Leica
For information 800-248-0123
www.leica-microsystems.com

Piezo Stage System

The NZ500CE NanoScanZ Piezo 500 micron stage system is an innovation in microscope automation designed for researchers using deconvolution and three-dimensional imaging techniques. The NanoScanZ Piezo 500 offers five-nanometer repeatability and closed-loop control with a subangstrom resolution piezo resistive sensor. The stage provides piezo speed with a sample holder area large enough for well plates in addition to standard slides and petri dishes. The 500-micron movement enables the NZ500CE to focus over an entire well plate without having to adjust the microscope focus mechanism. The stage complements the speed of the newest digital cameras and accomplishes in milliseconds what would take seconds for conventional rotary focus drives. By moving the sample instead of the objective, the NanoScanZ offers advantages over objective-based piezo systems, including faster movement and settling and the flexibility to create Z stacks with multiple objectives.

Prior Scientific
For information 781-878-8442
www.prior.com

PFilter Set Tailored to Nanocrystals

Invitrogen's Molecular Probes Qdot 625 nanocrystal fluorescent label is matched to work with the Brightline fluorescence filter set from Semrock. This new label-filter combination allows the ultimate in brilliant imaging or quantitation by combining Invitrogen's stable and technically advanced Qdot nanocrystal fluorescence labeling technology with Semrock's matched "no burn out" high performance fluorescence filters. Qdot fluorescence labeling products' imaging performance comes from their robust nanocrystal structure combined with a customizable surface. As one example of their wide range of applications, the Qdot 625 streptavidin conjugate fills an important spectral gap for dense multicolor imaging experiments and exhibits greater brightness than the longer-wavelength Qdot nanocrystals. Coupled with Semrock's high-brightness matched fluorescence filter set, these products are suitable for demanding experiments such as live-cell, single-molecule dynamic imaging. The Qdot labels are not only bright, but usable in applications that require long-term photostability.

Semrock and Invitrogen
For information 585-594-7003
www.semrock.com

Multispectral Imaging System FX

The Kodak In-Vivo Multispectral Imaging System FX is a new digital system for fluorescent molecular imaging that enables life science researchers to study molecular abnormalities that are the origin of disease at a very early stage. The system is designed to precisely locate and monitor changes in the molecular activity of specific cells or organs long before morphological changes can be detected, expediting the development of effective therapeutics for disease treatment. The system's unique multispectral tuning of excitation light greatly enhances image sensitivity, allowing for the identification and separation of multiple optical biological marker signals in subjects such as small animals, while removing background noise. The system's sophisticated software automatically generates and analyzes a series of images taken at different wavelengths with spatially coregistered X-ray and white light images for improved localization of biomarkers in vivo. The system's multimodal capabilities include bioluminescence and radioisotopic signal detection.

Carestream Health
For information 877-747-4357
www.carestreamhealth.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.